REPORT RESUMES

ED 010 763 SE 000 014

A STUDY OF LARGE GROUP-SMALL GROUP INSTRUCTION IN REGENTS CHEMISTRY COMPARED TO CONVENTIONAL INSTRUCTION.

BY- WINTER, STEPHEN S. AND OTHERS

STATE UNIV. OF N.Y. BUFFALO, COLL. AT BUFFALO
ROCHESTER UNIV., N.Y.

PUB DATE APR 65

EDRS PRICE MF-$0.18 HC-$5.12 128P.

DESCRIPTORS- *CHEMISTRY, *SECONDARY SCHOOL SCIENCE, *SMALL GROUP INSTRUCTION, *LARGE GROUP INSTRUCTION, ACHIEVEMENT, CONVENTIONAL INSTRUCTION, INSTRUCTION, BUFFALO, ROCHESTER

STUDENTS IN EXPERIMENTAL LARGE GROUP-SMALL GROUP CHEMISTRY CLASSES WERE COMPARED TO STUDENTS IN CONVENTIONALLY ORGANIZED CHEMISTRY CLASSES FOR GAIN IN KNOWLEDGE OF CHEMISTRY AND OTHER LEARNING OUTCOMES. THE 1,241 STUDENTS INVOLVED IN THE INVESTIGATION WERE ENROLLED IN 14 URBAN, SUBURBAN, AND RURAL SCHOOLS. EXPERIMENTAL CLASSES RANGED IN SIZE FROM 78 TO 153 STUDENTS FOR LARGE GROUP INSTRUCTION AND FROM 7 TO 32 FOR SMALL GROUP ACTIVITIES. DATA WERE OBTAINED FROM STUDENT RECORDS, THE REGENTS EXAMINATION, AND TESTS CONSTRUCTED FOR THE STUDY. STUDENTS WERE PRETESTED FOR KNOWLEDGE OF CHEMISTRY, SCIENCE INTEREST, SCIENCE REASONING AND UNDERSTANDING, AND SCHOLASTIC ABILITY. THEY WERE TESTED FOR ACHIEVEMENT AFTER EACH OF SIX UNITS AND AT THE END OF THE YEAR FOR SCIENCE INTERESTS, AND SCIENCE REASONING AND UNDERSTANDING. DIFFERENCES IN THE EFFECTIVENESS OF THE TWO PATTERNS OF INSTRUCTION WERE DETERMINED BY ANALYSIS OF COVARIANCE. REGRESSION ANALYSIS WAS USED TO RELATE SUCCESS OF DIFFERENT TYPES OF STUDENTS TO THE TWO PATTERNS OF INSTRUCTION. ACHIEVEMENT OF THE EXPERIMENTAL STUDENTS WAS SIGNIFICANTLY HIGHER THAN ACHIEVEMENT OF STUDENTS IN CONVENTIONAL GROUPS ON THE REGENTS EXAMINATION AND THE UNIT TESTS. (AG)

ED010763

A Study of
Large Group-Small Group Instruction in Regents Chemistry
Compared to
Conventional Instruction

STEPHEN S. WINTER and S. DAVID FARR
State University of New York at Buffalo

and

JOHN J. MONTEAN and JOHN A. SCHMITT
University of Rochester

SE 000 014

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

SEP 6 - 1966

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A STUDY OF
LARGE GROUP - SMALL GROUP INSTRUCTION IN REGENTS CHEMISTRY
COMPARED TO
CONVENTIONAL INSTRUCTION

BY

Stephen S. Winter and S. David Farr
State University of New York at Buffalo

and

John J. Montean and John A. Schmitt
University of Rochester

Supported by the State Education Department of New York State
and Fourteen Western New York School Districts

April, 1965

Educational Research Center
School of Education
State University of New York at Buffalo

April, 1965

FOREWORD

In addition to being a research undertaking in an important area of curriculum, the study of Large Group Instruction in Chemistry represents an innovation in the support and design of educational research. It is a successful demonstration of a thesis, vigorously encouraged by Dr. Lorne H. Woollatt, Associate Commissioner of Education for Research and Evaluation, State Education Department of New York, that cooperative educational research can produce more significant and more far-reaching results than research limited to single schools and single investigators. To be exact, the study shows that several schools with comparable programs provide a more diversified testing ground for educational research than does a single school; that university professors trained in research methodology make a contribution to curriculum research which can not easily be matched by school personnel alone; and that given cooperative support, both financial and moral, enough patience and good-will, and a good amount of just plain effort, a large scale project can obtain results which elude the less sensitive designs of smaller studies.

Exactly where the idea for this cooperative curriculum research project arose is unknown to the editor of this report because he joined the project after some important early decisions had been made. (And until this day of reckoning, when the report is finally being written, he was more concerned with the present and the future of the research than with its past.) His apologies to those whom he might slight through ignorance. When he joined the project, the study of Large Group - Small Group Instruction in Chemistry Compared to Conventional

Instruction had been agreed upon, Drs. Gerald T. Kowitz and Henry Hausdorff of the State Education Department had assembled a group of twelve school systems willing to support it, and Dr. John D. Scheller, of Amherst Central, had agreed to serve as the coordinator for their contribution. The principal investigators were asked to submit a proposal for a research design which, after careful review, met the approval of the Department. Thus, with the leadership of Dr. Scheller and much help from Drs. Kowitz and Hausdorff, the study got off to the happy beginnings which it was able to maintain to the very end. Perhaps the only unfortunate note that crept in during the three years of work was the realization, as datum followed datum, as punch-card box was stacked on top of punch-card box, and as the computer printed and printed and printed, that in our inexperience with a study of this magnitude we had underestimated the wealth of material at hand. We had promised too little, too soon. This report makes only restricted uses of the data; those that were fully planned in advance to answer the questions posed in the initial proposal.

The principal investigators for the study were Drs. John J. Montean and John A. Schmitt of the University of Rochester and Drs. S. David Farr and Stephen S. Winter of State University of New York at Buffalo. Dr. Winter also served as Project Coordinator and as editor of this report. Drs. Montean and Winter had major responsibilities for the chemistry-related problems of the research; Drs. Farr and Schmitt were chiefly concerned with evaluation, research design and analysis. In all this, they depended heavily for ideas and personal evaluations on the teachers in whose classrooms the study was carried out:

Mr. George Martiny	Akron Central High School
Miss Esther B. Tomlinson	Alden Central High School
Mr. Elton W. Petersen	Amherst Central Senior High School

Dr. Samuel W. Bloom	Benjamin Franklin High School
Mr. Harold L. Kruger	Brighton High School
Mr. Lewis A. McCall	Canandaigua Academy
Mr. Arthur H. Root	Clarence Central High School
Mr. Ben Varco	Eden Central High School
Mr. Leonard S. Blake	Kenmore East High School
Mr. Lawrence Gideon	Orchard Park Central High School
Mr. Albert A. Deney	West Seneca Central High School
Mr. Frank J. Tuzzolino	Williamsville Senior High School

These teachers contributed items for the tests and provided tables of specification for them; they tried the instruments and criticized them; and always, they were willing to work sudden and difficult requests into their teaching schedules. In the final year of the study, when the pace became more rapid but also somewhat more predictable, they were joined by:

Mr. Leonard Weiss	Cleveland Hill High School
Miss Beatrice Elye	Cleveland Hill High School
Mrs. Camille Gilmour	Hamburg Central High School

To all of them the principal investigators express their thanks, and also their admiration for the good nature and good will maintained consistently throughout leisure and chaos alike.

The study owes much to four research assistants, all graduate students at State University of New York at Buffalo. Messrs. James Blaydon, Jack Dudley, Richard Egelston, and William Martin likewise refused to get rattled or upset or even admit to being tired when testing from 7 A.M. until 4 P.M. at schools miles apart, when sitting in front of computers that were unable to stand the heat and humidity of a midsummer weekend midnight and failed to compute as scheduled, or when the papers piled high on their desks threatened to crowd them out of their rooms. As representatives of the study office who most frequently visited the schools, they helped maintain the happy communications.

Much help was given also by a number of other persons: Mrs. Gail

Hofmann and Mrs. Sandra Yarnes, our secretaries; the guidance counselors, principals, and their secretaries at the schools; Mr. Howard English, of the computer center, Mr. Robert C. Kochersberger. To all of them our thanks.

For financial support of the study we are indebted to the State Education Department and the Boards of Education of the participating school districts. The Department, and the chief school officers of the participating school districts:

Mr. Edward E. Allen	Supervising Principal	Akron Central H. S.
Dr. Wilson R. Conrad	Supervising Principal	Alden Central School
Dr. John D. Scheller	Supervising Principal	Amherst Central Sr. H. S.
Mr. Fred B. Painter	Supt. of Schools	Brighton H. S.
Mr. Robert Helmer	Supt. of Schools	Canandaigua Academy
Dr. Arthur Shedd	Supt. of Schools	Clarence Central H. S.
Mr. Walter J. Heffley	District Principal	Cleveland Hill H. S.
Mr. Edwin C. Peck**	District Principal	Eden Central School
Dr. Richard Burau	District Principal	Eden Central School
Mr. Harry H. Hatten	Superintendent	Hamburg Central School
Dr. Carl W. Baisch**	Supt. of Schools	Kenmore Public Schools
Mr. C. Sherwood Miller	Supt. of Schools	Kenmore Public Schools
Mr. Elmer D. Handel	District Principal	Orchard Park Central H. S.
Dr. Robert Springer*	Supt. of Schools	Rochester Public Schools
Mr. Herman Goldberg	Supt. of Schools	Rochester Public Schools
Mr. Alfred W. Goodreds	Supervising Principal	West Seneca Central School
Dr. William E. Keller	Supervising Principal	Williamsville H. S.

* deceased ** retired

provided much more than funds, however, through their constant interest and encouragement.

And finally, to the students who so willingly participated in this experiment, who subjected themselves cheerfully, if not voluntarily, to several three-hour blocks of solid testing, our most sincere appreciation. We hope they feel that their cooperation will lead to better education for their younger brothers and sisters, and for their children.

This report has been written by all four principal investigators. As editor, however, I cannot disclaim major responsibility for inadvertent shortcomings.

Buffalo, New York
1965

Stephen S. Winter

TABLE OF CONTENTS

Page

FOREWORD iii

LIST OF TABLES ix

Chapter 1 - INTRODUCTION 1

Chapter 2 - DESIGN AND INSTRUMENTS 7

Chapter 3 - THE STUDY 23

Chapter 4 - PUPIL ACHIEVEMENT 36

Chapter 5 - PREDICTING ACHIEVEMENT 75

Chapter 6 - INSTRUCTIONAL PROCESSES 90

Chapter 7 - COST 100

Chapter 8 - RECOMMENDATIONS 107

Chapter 9 - A NOTE ON PROCEDURE 111

REFERENCES 115

PRECEDING PAGE BLANK-NOT FILMED

LIST OF TABLES

Table		Page
I	Tables of Specification for the Unit Tests	13
II	Distribution of Item-Types Among Unit Tests	18
III	Schedule of Test Administrations	28
IV	Characteristics of Conventional Schools	31
V	Characteristics of Large Group - Small Group Schools . . .	32
VI	Comparison of Mean Scores on Eight Control Variables . .	35
VII	Correlation Coefficients Between Eight Control Variables .	38
VIII-a	Analysis of Covariance with Part One of the June, 1964, Regents Examination	40
VIII-b	Analysis of Covariance with Part Two of the June, 1964, Regents Examination	41
VIII-c	Analysis of Covariance with Total Scores on the June, 1964, Regents Examination	42
IX-a	Analysis of Covariance with Scores on the Unit Test Covering the Periodic Table of Elements	44
IX-b	Analysis of Covariance with Scores on the Unit Test Covering Water, Solutions and Ionization	45

Table Page

IX-c Analysis of Covariance with Scores on the Unit Test Covering the Halogen Elements 46

IX-d Analysis of Covariance with Scores on the Unit Test Covering the Sulfur-Nitrogen Families 47

IX-e Analysis of Covariance with Scores on the Unit Test Covering the Principles of Chemical Reactions 48

IX-f Analysis of Covariance with Scores on the Unit Test Covering Organic Chemistry 49

X-a Analysis of Covariance with Total Scores on Type One Items 51

X-b Analysis of Covariance with Total Scores on Type Two Items 52

X-c Analysis of Covariance with Total Scores on Type Three Items 53

X-d Analysis of Covariance with Total Scores on Type Four Items 54

XI Analysis of Covariance with Science Reasoning and Understanding Posttest 57

XII Comparison of Scores on Test of Science Reasoning and Understanding Taken Before and After the Year of Instruction 58

XIII Comparison of Scores on Kuder Preference Record, Science Scale, Taken Before and After the Year of Instruction 60

XIV Analysis of Covariance with Kuder Preference Record Posttest 61

Table Page

XV-a Comparison of Sample Selected for Covariance Analysis with Total Students - Large Group Instruction 65

XV-b Comparison of Sample Selected for Covariance Analysis with Total Students - Conventional Instruction 66

XVI Comparison of Uncorrected Scores of Large Group Students with Conventional Students 67

XVII Number of Students Who Took Tests in September, October and June 69

XVIII Intercorrelation Matrix Between Recall, Comprehension, Application, and Higher Order Competency Items 72

XIX Correlation Coefficients Between Item Types and Other Measures 73

XX Regression Equations Between the Chemistry Pretest and the Unit Tests and the 1964 Chemistry Regents Examination 77

XXI Regents Examination Scores Associated with Specific Test Scores 79

XXII Multiple Regression Equations Between Predictor Scores and the 1964 Chemistry Regents Examination 81

XXIII Behavior Categories 91

XXIV Objective Categories 94

XXV Comparison of Grouped Teacher Behaviors 97

XXVI Adjusted Teacher-Pupil Loads and Teacher Preparation Time 105

Chapter 1

INTRODUCTION

Large group instruction is a relatively recent phenomenon in American secondary education. Its origins are not entirely clear from the literature; however, there is little doubt that the work of the Commission on the Experimental Study of the Utilization of Staff in the Secondary School, whose director was J. Lloyd Trump, gave impetus to the spread of this pattern of instruction. The Commission's report, GUIDE TO BETTER SCHOOLS (Trump and Baynham, 1961), gave the detailed rationale for the adoption of methods of instruction which brought together groups of 300 pupils for purposes that could be achieved as well with groups of that size and complemented this with smaller groups that ranged in size down to supervised instruction of one pupil at a time for purposes in which the sole learner had to be segregated. Among the objectives of the report seems to have been an attempt to break the habitual 25-30 pupils per class which has stifled flexibility in the deployment of instructional resources in American secondary schools.

While the Trump report drew upon selected explorations with large and small instructional groups, at the time of its publication few empirical studies of the results of such changed practices in the secondary schools had been completed. What little evidence existed was chiefly in the form of testimonials or of uncontrolled measurements of achievement. Such evidence is suspect, and far short of desirable. However, these relatively unsophisticated studies invariably were favorable to the large group - small group organization on one or another criterion. Thus there was good indication that large group instruction did not hurt youngsters' achievement of content; that it was pleasant to them, or at least not traumatic; and that it was viewed favorably also by the teachers who tried it. For instance, Lisonbee (1962) reported the following personal observations concerning a large group class in chemistry: discipline was no problem and rapport between instructor and students was excellent throughout; student participation increased as time went on and it became possible to have class discussions; personal help could be given to students in small groups; 95% of the students liked the course, although "90% voted equally for the large class and a regular one." Finally, in the important area of achievement, "it was indicated that the experimental group held their own and showed a probability of doing slightly better than expected." A dissertation by Breedlove (1963) generally confirmed these results adding that large group students gain greater understanding of the methods of science.

Lisonbee and Breedlove's studies deal specifically with chemistry, the subject of the present investigation. They are fairly typical of other reports, however. Thus Johnson and Lobb (1961) report on the effect of class size on the "achievement, attitudes, and behavior of the learners" in Jefferson County, Colorado:

> As a result of this study it was determined that the size of class did not in itself make any significant difference. Specifically, the experiment produced these findings; first, there was no significant difference in the achievement of pupils in classes of 20, 30, 60, and 70; second, small groups of high capacity learners were not academically or economically feasible; and further, students had not been harmed by participating in large group work.

Other research of similar nature, with the same kinds of results, has been reported in the annual issue of the Bulletin of the National Association of

Secondary School Principals devoted to team teaching and large group instruction.

At the college level, where the lecture has been the traditional method of instruction, more extensive investigations of large group instruction have been conducted. While some of these were somewhat more adequate in design than most high school studies, an exhaustive review of the literature by Fleming (1958) came up with the so common, yet so discouraging, result in educational research: no significant difference. To be exact:

> Once again, therefore, it has to be said . . . that class size itself does not appear to be a significant factor in achievement, and further support is lent to the (idea) . . . that the secret difference between the success of one teacher must be looked for in something other than the mere size of the group under tuition.

Since 1961, additional reports have come in of the effectiveness of large group instruction. In general, they reinforce the conclusions which were evident when this study was planned in the fall of 1961.

The Advantages of Large Group Instruction

While large group instruction had not proved itself advantageous--nor disadvantageous--in terms of achievement, there are a number of other factors to be considered. These are given by Trump and Baynham (1961, p. 30-31):

> In tomorrow's schools, more students will be exposed to skilled teaching in all subjects because the most capable and experienced teachers in specific fields will teach large classes. Every teacher, inevitably, is more experienced in one subject or one phase of a subject than in another. So the students can be better motivated by contact with the very best teacher available for that phase of the subject. The large class will avoid the duplication of effort required when teachers must teach the same subject matter to a number of

> classes, as in today's schools.
>
> Tomorrow's schools will find a number of other advantages in large class instruction. It makes technological equipment more economically feasible. Many schools which can't afford these aids for five or six customary-size classes will be able to equip a single large classroom.
>
> Economy--in this case economy of time--will allow tomorrow's schools to schedule more presentations by outside specialists, by teachers from nearby universities who are experts in certain fields, and by community resource persons. For example, in one of the experimental schools, a local physician came to the school and talked at one time to all of the biology students about the circulation of blood. His experiences as a surgeon gave him more background about the heart and blood system than any teacher in the school possessed. The crowded days of most of these people make it difficult for them to leave their own work for more than an hour or two at a time.
>
> For students, large classes will offer another particularly important advantage: They will serve as transitional experiences for college classes and for many other occasions of adult life. Students can learn to take notes, hold back questions until an appropriate time, and develop more responsibility for planning their own learning.

Rationale for This Study

The advantages cited by Trump and Baynham are indicated by logic and supported by the testimonials which had been published at the time this study was planned. The principal investigators in this research felt, however, that through an adequate plan for research, some of the indications could be corroborated by firm evidence and some questions which had not been raised before could be explored as well. To be exact, it was felt that a massive and carefully planned research effort could go beyond the gross measures of achievement which had typically resulted in "no significant difference" to explore relationships between the competence of the learner, the method of instruction, and the cognitive outcome achieved (Bloom, 1956). Secondly, it was thought that questions relating to affective learnings--attitudes, motivations--which had been explored chiefly in the college setting and for which differences were recorded (Gerberich and Warner, 1936; Ruja, 1954; Bloom, 1953) could also be explored, and through more sophisticated approaches than the direct questionnaire. Finally, the claims made for equal or lesser cost under the large group - small group pattern while adding to the teacher's preparation time were felt in need of empirical verification. Thus the study was planned to supplement what had been reported before and to seek evidence in aspects of the large group organization which heretofore had been unresearched.

A last dimension which required empirical study was the character of large group - small group instruction. The term "large group - small group instruction" is too imprecise to serve as an adequate description of the learning situation. More accurate descriptions of the teaching tactics used both under the conventional and under the large group - small group pattern were felt necessary to explain any differences than might be discovered. Consequently, the study included an extensive exploration of the methods used in each of the classrooms so that learning interactions as well as group size could be examined as correlates of achievement or motivational results.

Time Schedule

The study was begun late in the fall of 1961 and planned for completion in 1964. Fortunately, no major deviations were required until the very end when a number of factors delayed data analysis and evaluation, and the production of the final report.

During 1961-62, it was planned to identify the instruments needed for the study and to begin their construction, testing, and evaluation. The year was also used to develop the communications processes needed for this study which involved teachers at 14 schools and two universities. 1962-63 was set aside to complete instrument production and to test research methodology and procedures with a fraction of the schools and a small block of chemistry content and teaching time. Finally, the complete study was carried out during the entire school year 1963-64, beginning with a teacher administered pretest on the first day of instruction and ending with the Regents Examination late in June.

Chapter 2

DESIGN AND INSTRUMENTS

The empirical investigation of the results of large group - small group instruction in Regents chemistry as compared to conventional instruction set itself the following goals:

1. To determine which organizational pattern produced greater achievement after allowance had been made for differences among the pupils through an analysis of covariance treatment:
   a. In content as measured by the Regents Examination.
   b. In content as measured on tests of six common units.
   c. In content as differentiated according to a classification scheme of competencies in the cognitive domain following that of Bloom (1956).
   d. In science interest.
   e. In science reasoning and understanding.

2. To identify the kind of learners for whom the above achievement statements might be true. This part of the study attempted to answer questions such as: Do girls achieve more highly on memorative content materials when taught in large groups? Do students with greater critical thinking ability change toward greater science interest when taught by conventional organization? Do students with a poorer academic record achieve more highly in one or the other organizational pattern?

3. To ascertain whether organizational pattern also identifies instructional procedures, or whether instructional tactics with each pattern

vary so much that teacher characteristics are more important than organizational pattern.

4. To assess whether there are cost differentials that might preclude adoption of the large group pattern--even if it proved educationally desirable--except in the most highly supported school districts.

The design of the study, therefore, required a large number of measurement instruments some of which were available among the published tests, some of which, however, had to be produced for the study.

Criterion Instruments

The following tests were used as criterion instruments:

1. The Regents Chemistry Examination for 1964, to measure achievement in accordance with the Regents Chemistry Syllabus.

2. The Kuder Preference Record, Vocational Form CM, which has a science interest scale, to measure changes in science interest.

3. The Test of Science Reasoning and Understanding produced by Dressel for the Cooperative Study of Evaluation in General Education of the American Council on Education to measure changes in science reasoning and understanding.

These three instruments were used as available. In addition, achievement on six units of the Regents chemistry course was determined by six unit tests devised specifically for this study. Moreover, the items for these tests were constructed in such a manner that they represented the various categories of competencies in the cognitive domain so that the unit tests served also as measures of the different intellectual competencies.

Control Instruments

In order to characterize students and to segregate them into educationally meaningful groups, the study employed a large number of instruments covering attributes presumed independently influential in determining success on the criterion variables. Among the dimensions explored, most of which subsequently were shown to result in relatively independent scores, were intellectual competence and achievement; science interests; prior knowledge of chemistry and chemistry aptitude; study habits; critical thinking; science reasoning; and some socio-economic factors. Most instruments needed for these purposes were available among published tests:

1. The Holzinger-Crowder Unifactor Tests, which have four parts (verbal, spatial, numerical, and reasoning), characterized the competencies usually associated with IQ.

2. Students' grades in academic courses from grade 9 onward, obtained directly from records on file in the schools, characterized their general scholastic achievement.

3. Students' science interests were characterized by the Kuder Preference Record administered as a pretest (as well as a posttest. See above).

4. Chemistry aptitude was characterized by the Iowa Placement Examinations, Chemistry Aptitude, Form M. Since this test was published in 1942, its part 4, dealing with chemical facts, was replaced by a Chemistry Achievement Pretest constructed especially for this study.

5. Study habits of the students were characterized by the Brown-Holtzman Survey of Study Habits and Attitudes.

6. Critical thinking was characterized by the Test of Critical Thinking, Form G, developed by the Cooperative Study of Evaluation in General Education

of the American Council on Education.

7. Science Reasoning Ability was characterized by the Test of Science Reasoning and Understanding (See above) given as a pretest.

8. Socio-economic factors were characterized by information supplied by the subjects of this study regarding their parents' education and occupational status.

It was found that not all of these instruments gave information regarding student competences which was sufficiently independent of the other instruments to serve as a useful characterization. The Iowa Placement Examination, The Survey of Study Habits and Attitudes, the Test of Critical Thinking and the socio-economic data were not subsequently used in the analysis. An additional characterization, of course, was sex of the student.

Other Variables

The achievement of the third goal of this study, the description of teaching tactics used by each of the teachers throughout the year, required the construction of still another instrument. The literature contains plentiful suggestion concerning the construction of check lists of teacher performance. These, however, have supervisory and evaluative purposes which were contrary to the objectives of the description instrument needed for this study. On the other hand, Flanders has devised a method for describing the classroom behaviors of teachers (Flanders, 1960). His method, however, requires an observer. This eliminated the Flanders instrument. A measure was developed for this study on which each teacher recorded his behavior in categories important to the description of teaching tactics and which he could fill in daily with relatively little difficulty.

Finally, an analysis of teacher time use and space use served as the basis for the determination of teaching costs of the two programs. Professor Milton Pullen of the University of Rochester served as consultant to this part of the study. The cost analysis went, in part, beyond the participating schools in order to extend the sample.

The Unit Tests

At one of the first meetings of groups involved in this study, principal investigators and teachers, it was ascertained that each teacher teaches as a discrete unit the material in the Regents Syllabus related to each of the following topics:

1. The Periodic Table and the Properties of Atoms.
2. Water, Solutions, and Ionization.
3. Principles of Reaction.*
4. The Halogen Elements.
5. Elements of the Sulfur and Nitrogen Families.
6. Organic Chemistry.

*Subsequently it was discovered that Principles of Reaction were not taught as a unit but merely reviewed as such. See Chapter 3.

These six were therefore chosen as the units which could give partial information on student achievement and for which tests could be developed with questions that contained items written specifically in accordance with all categories of Bloom's taxonomy (1956). Teachers were invited to submit old tests on these units to the principal investigators to serve as the source of items for the unit tests. With these tests and the Regents Syllabus as guides, tables of specifications were prepared for each of the unit tests. These were submitted to the teachers for an indication of the relative emphasis on each of the topics within the unit tests. Responses were averaged for the final distribution of items in accordance with the table of specifications. The tables of specification for the unit tests, with an indication of the relative emphasis within each test, are given in Table I.

The principal investigators then wrote about 100 items of multiple choice type for each unit. The items were adapted from the items submitted by the teachers and roughly approximated in numbers the distributions within the tables of specification. Particular care was taken to include items which could be rated according to each of the intellectual competencies of Bloom.

Several compromises had to be made in adapting Bloom's taxonomy to the study. In the first place, the taxonomy is quite detailed. It was recognized that only the major subdivisions of the taxonomy could be represented by a sufficient number of items to provide an adequate sample of that competence among the tests. Secondly, it was found that items requiring evaluation, analysis, or synthesis are extremely difficult to construct within the multiple choice style required for the reliable handling of the data in this study. Thirdly, Bloom's definitions of categories are relatively cumbersome and at times unclear. A simpler scheme for classifications was required. Accordingly, the following four categories were established for the purpose of this study:

Type I. Recall: Any item which had been taught in substantially identical form, and which required merely the recall of that bit of information. Rephrasing, inversion of sentences, and similar changes of form were considered not to change an item from the recall category. Examples of recall items, from several of the tests, are:

Which of the following water solutions does not conduct electric current?

a) $CuSO$
b) $Al_2(SO_4)_3$
c) HCl
d) C_2H_5OH
e) KNO_3

TABLE I

TABLES OF SPECIFICATION FOR THE UNIT TESTS

Periodic Table

1.	History	4%
2.	Principal sub-atomic particles, including atomic structure, atomic number, atomic weight, isotopes	17%
3.	Electron levels (orbitals and spectra)	10%
4.	Bonding--ionic and covalent bonds, characteristics of compounds, electron-dot diagrams of simple compounds	15%
5.	Valence, metals, non-metals, activity and electronic structure, activity and size	20%
6.	Periodic Table, families, periodic variation, other relationships	16%
7.	Symbols, formula writing, molecular weight from formula	16%

Water, Solutions, and Ionization

1.	Solutions--definitions, rate of solution, solubility, saturation, solvents	15%
2.	Water--crystallization, water of crystallization, dehydration, purification and distillation	10%
3.	Theory of ionization, water solutions of electrolytes, molten electrolytes, non-electrolytes, electrolysis	25%
4.	Concentration (molar) and other	18%
5.	Boiling point, freezing point changes, vapor pressure	7%
6.	Acids, bases, neutralizing action, pH, hydrolysis	25%

Principles of Reaction

1.	Factors causing completion of reaction (precipitate, gas, oxidation, etc.)	20%
2.	Activity of metals, displacement series; displacement of some non-metals	16%
3.	Redox, agents, oxidation number	20%
4.	Energy changes, stability, exo-endothermic	9%
5.	Types of reactions	16%
6.	Equilibrium and Le Chatelier's Principle, Common Ion Effect (no numerical problems)	11%
7.	Velocity of Chemical Change and factors affecting it, catalysts	6%

TABLE I

TABLES OF SPECIFICATION FOR THE UNIT TESTS
(Continued)

The Halogen Elements

1.	Family in periodic table, changes of properties in the family	12%
2.	Occurrence and production of elements	12%
3.	Properties and reactions of the elements	23%
4.	HX compounds, production, uses; MX salts	17%
5.	Oxy-acids, names, oxidation numbers, etc.	12%
6.	Redox in the halogen family and review of redox	13%
7.	Review of weight, volume, composition problems	10%

Sulfur and Nitrogen Families

1.	Sulfur, forms, Frasch, reactions	7%
2.	SO_2 and H_2SO_3, preparation and reactions	11%
3.	H_2S, preparations, reactions, acidity, test for	12%
4.	$H_2SO_4 - SO_3$, reactions, identification, Contact process	11%
5.	N_2, sources, uses in combined form	7%
6.	N-fixation, including Haber, Ostwald	8%
7.	Chemical equilibrium, review	8%
8.	HNO_3--properties, preparation, reactions	13%
9.	NH_3--properties, preparation, reactions, weak base	12%
10.	Oxides of N	7%
11.	Other elements in family (P, etc.)	5%

Organic Chemistry

1.	Definitions, tetravalent C, difficulty of reactions, etc.	7%
2.	Homologons series, series formulas, names	16%
3.	Structural formulas and isomerism	14%
4.	Unsaturation, saturation, and simple reactions--double, triple bonds	14%
5.	Benzene, simple reaction, simple derivatives	6%
6.	Functional groups (OH, RCOOH, RCOOR)	16%
7.	Ethers, aldehydes	6%
8.	Organic fuels--petroleum, gasoline, octane, etc.	10%
9.	Polymerization, and other reactions	7%
10.	Other (biologic applications, plastics, etc.)	4%

Which of the following is a good reducing agent?

a) S b) Br_2 c) O_2 d) C e) F_2

Type II. Comprehension: Any item which required the application of a principle under circumstances not identical with the context in which the principle had been taught, but in such form that the necessary principle is implied in the question. Examples of comprehension items are:

The ion in which chlorine has an oxidation number of -1 is

a) Cl^- b) ClO^- c) ClO_2^- d) ClO_3^- e) ClO_4^-

The atomic number of a certain element is 53. An atom of this element must contain

a) 53 neutrons
b) 26 neutrons and 27 protons
c) 1 neutron, 26 protons and 26 electrons
d) 53 protons
e) 53 particles of all kinds

Type III. Application: Items as above except that the principle needed is not implied in the question, and the student must choose the needed principle as well as apply it for the solution. Most numerical problems were considered in this category. Examples are:

The reaction between $Ba(OH)_2$ - HNO_3

a) approaches completion because a precipitate is formed.
b) approaches completion because a gas is formed.
c) approaches completion because an un-ionized substance is formed.
d) approaches completion because two of the above both occur.
e) does not approach completion; none of the above applies.

Al_2O_3 and CCl_4 are the correct formulas of the oxide of aluminum and the chloride of carbon. The formula of the compound aluminum carbide is therefore

a) Al_4C_2 b) Al_3C_4 c) Al_4C d) Al_4C_3 e) Al_2C_3

Type IV. Higher Competencies: This category included items which required analysis of a complex situation and the subsequent drawing of analytical, synthetic, or evaluative inferences. Examples of this kind of item are:

The following equations describe the reaction of Fe with HCl and with Cl_2

$$Fe + 2HCl \longrightarrow FeCl_2 + H_2$$

$$2Fe + 3Cl_2 \longrightarrow 2FeCl_3$$

This proves that

- a) Cl_2 is a good reducing agent
- b) HCl is a good reducing agent
- c) HCl is a good oxidizing agent
- d) Cl_2 is a better oxidizing agent than HCl
- e) Cl_2 is a better reducing agent than HCl

As a 6N solution of sulfuric acid is diluted to 0.1N, the pH

- a) increases
- b) decreases
- c) remains the same
- d) may increase or decrease
- e) cannot predict the answer

It was recognized in the construction of these items, as Bloom already pointed out, that all items require an initial recall of information so that, in this sense, they depend initially upon rote memory. The distinction is between memory alone and memory plus other intellectual behaviors. Secondly, the distinction between categories is heavily dependent on judgments on which consensus is not perfect. Thirdly, any item, whatever the ingenuity in its construction, can be turned into a recall item if the teacher chooses to teach its content in that manner. Nevertheless, the length of the sequence of information which has to be memorized differs for the various items and, even in that case, the categories distinguish at least among different complexities in the material memorized.

Each of the items for each test was identified according to the above type by concensus between Drs. Montean and Winter after tabulation of ratings given the items by the collaborating chemistry teachers. The items were then split into two tests, one typically somewhat long but considered "easy" by the students who took it and containing chiefly Type I items--sometimes supplemented by a few Type II items--and the other containing all other items. As each participating teacher completed the unit for which the unit test had been constructed, he was given enough copies of each of the two tests for half of his classes. He was not told which test was "easy" and which "hard" but asked to distribute the tests among his pupils without any specific ordering. Teachers were asked to use the results from these tests in their grading to insure usual motivation and were promised, and subsequently sent, normative and scaling data for each of the two tests.

The tests were returned to Buffalo for grading and item analysis. Item difficulties, item discrimination, and point-biserial correlation values were computed (Gulliksen, 1950). All items were then reviewed in a meeting of all teachers with the principal investigators to explore difficulties that might have arisen during administration; problems with comprehension, applicability, style of response, etc. Only items which were agreed upon by all teachers, which discriminated significantly, and were answered correctly by approximately 40-80% of the students were retained. From this acceptable group of items, a final 50-item test was constructed for the study, with items distributed according to the table of specifications and containing approximately 40% of Type I items and 20% of items of each of the other types. It was difficult to adhere to both sets of restrictions, and distribution of item type was compromised in the final construction of tests (Table II). Two forms of each test were produced with identical items arranged in a different order.

The Chemistry Pretest

The Chemistry Pretest was produced in essentially the same fashion as the unit tests. Since it was designed to measure prior chemistry knowledge, items were chiefly of Type I. Their distribution followed the topics of the Regents Syllabus in chemistry. Items were written by the principal investigators, sent to the schools for administration during the first week of

TABLE II

DISTRIBUTION OF ITEM-TYPES AMONG UNIT TESTS

TEST NAME	NUMBER OF ITEMS OF TYPE			
	I	II	III	IV
Periodic Table	15	12	10	13
Water, Solutions, and Ionization	17	15	12	6
Principles of Reaction	12	14	18	6
Halogens	15	18	8	9
Sulfur and Nitrogen	22	10	8	10
Organic Chemistry	14	30	5	1
	95	99	61	45

the school year 1962-63, discussed in meeting, and analyzed. From the result of the trial administration, a 40-item test of items with the required response characteristics and following the distribution of the Regents Syllabus was constructed and called the Chemistry Achievement Pretest. This test also was produced in two forms differing only in the sequence of the items.

The Instructional Process Report Form

The major purpose of the Instructional Process Report Form was to establish whether group size also implied a difference in the style of teaching. Consequently, the emphasis in the instrument was on the recording of the kinds of activities which the students experienced in each classroom, with an indication whether the objectives of the instructional activities differed in the two organizational patterns.

The instrument consisted of a series of simple response categories giving time, place, size of group, subject matter covered, kind of session, length of session, kind of homework, etc. Then, there followed an open-response section in which teachers recorded, also by code, the specific kinds of interactions that took place, whether teacher dominated, teacher-pupil exchange, pupil dominated. Also recorded was the planned purpose of the interactions and their duration. There were 23 categories under classroom interactions and 10 under objectives or purpose. In addition, the form contained space for remarks or aspects not provided for among the encoded categories.

This instrument, which proved quite functional when appropriately used, was developed through a series of trials of short duration followed by analysis of responses and discussion of results with the teachers who had to use it. For each trial, the principal investigators prepared a direction booklet for the completion of the forms, with the stipulation that the forms would be used for one section of Regents chemistry in all its scheduled activities--large group, small group, and laboratory; or conventional class meeting and laboratory--and in all non-scheduled contacts with students which concerned the course. Among the latter were pupil-solicited or teacher-suggested help to individuals or small groups; make-up work; help with projects; and other similar non-scheduled contacts. Typically, each teacher used about three report forms daily, one for classroom and two for other teaching-related activities or contacts. The instrument went through three two-week trials followed by careful revisions before categories were considered adequately exclusive and the response form adequately simple to allow rapid and unambiguous description of the teaching style. It was then used in the pilot study of approximately six weeks duration in which it proved satisfactory for extended use.

Pupil Achievement and Socio-economic Data

The gathering of data of pupil's past achievement, surprisingly, proved more difficult than anticipated. There is great diversity in the way schools maintain their records. Consequently, it was necessary to provide the guidance office staff in each school with considerable detail regarding the needed information before it could be extracted from the files without undue difficulty. Through the experience gained in obtaining achievement and socio-economic data for the pilot study, it was possible to design a form that met the needs although strange course titles or occasional inversions in the order in which the courses were offered still presented some difficulties. The form that was finally used allowed each student to check the courses he had taken in each academic area and in each grade from grade 9 on. Grades were later supplied for the checked courses by the guidance officials or their secretaries. Also on the form, the students recorded parents' occupation, parents' education, and personal vocational objectives.

Cost

The assignment of specific costs to Regents chemistry in the two organizational patterns proved to be the most difficult task confronting the study. School budgets are not kept in a manner that allows the easy isolation of costs attributable to one course. After thorough discussion of these matters with school business officials and principals, Professor Pullen, the consultant for this part of the study, recommended that the cost study limit itself to determination of teacher time costs, space costs, and recorded supervisory and developmental costs. While this eliminated a number of cost categories from consideration because data were not attainable, it was felt that this simplification of the study was justified because salaries make up the major fraction of the school budget, space costs also rank high, so that only lesser cost factors were neglected. Moreover, no logical basis was found for the assumption that laboratory or supply costs should differ significantly under the two plans. The ignoring of such factors meant, at worst, the ignoring of small differences among small budget factors.

Periods assigned to chemistry instruction rather than teacher pay was used as the basis for comparing teaching costs because the salary scales in the participating districts were not strictly comparable and, furthermore, the exact cost of instruction varies with factors that are not related to large group-small group or to conventional instruction. Hence, it was considered more important to know that one teacher instructed a total of 125 pupils per week than that a salary item of $7,000 performed such instruction.

Ultimately, a very simple form was devised to attribute teacher time to Regents chemistry and to make corrections for the fact that the number of exposures to chemistry could be either six or seven periods per week for both kinds of schools. In addition to teacher-pupil ratios, the form recorded spaces utilized and recent direct expenditures for developmental and supervisory costs. In order to extend the sample, a number of schools not otherwise in the study were added to obtain more broadly representative data for the cost analysis.

Chapter 3

THE STUDY

While a total of three years was required to complete the study of large group - small group instruction in Regents chemistry compared to conventional instruction, the data for the final comparisons were obtained during the year 1963-64 only. All earlier work was preparatory and was concerned with selecting and developing the measurement instruments needed for the study, with testing the efficacy of the instruments, with establishing the communications processes needed to collect the large amount of data easily and on time, and with testing the data reduction capabilities of the computers with the available computing programs. Since these preparatory activities had little direct bearing on the conclusions of the study, they will be treated very briefly.

1961-62

The first year was devoted only to instrument selection and instrument construction. It was found, as has been recognized elsewhere (Watson and Cooley, 1960), that there is a severe paucity of good instruments for measuring interests, motivations, and critical abilities in science. Moreover, there are no tests with validated items specifically categorized according to cognitive difficulty. Thus, the construction of unit tests with items following Bloom's taxonomy was immediately planned, and a later decision to construct a special Chemistry Achievement Pretest was based on the same reason.

During the first year, work was begun on all six unit tests, and the tests on Halogens, Sulfur-Nitrogen, and Organic Chemistry went through pretesting, analysis, and evaluation. The other tests could not be pretested because the appropriate units had been taught before the trial tests could be constructed and had to be carried over into the second year of preparation. The Instructional Process Report Form went through two brief trials and revisions. The other instruments were drafted, and a detailed work schedule was planned for the subsequent years.

1962-63 -- Pilot Study

The major activity of 1962-63 was the planning and execution of a six-week pilot run of the study. It was restricted to a sample of six schools (one later had to be dropped) and to the unit test on Sulfur-Nitrogen only, but otherwise followed in complete detail all the anticipated procedures of the full study. Thus, pretesting with the control instruments was carried out during the first weeks of school; students' past achievement data was gathered as needed; during the teaching of the unit on Sulfur-Nitrogen, which took about six weeks, the Instructional Process Report Form was completed daily; and final Regents Examination grades were collected. The data were then analyzed during the summer, although the analysis was limited to those aspects which needed to be checked to assure the success of the final study.

There were two important purposes to the pilot run. It was, in the first instance, a test of the data gathering and data handling procedures contemplated. As a result of pilot run experience, for instance, the pretest schedule was changed from two 3-1/2-hour blocks of time on successive days during the first or second week of school to a pretest period of 3-1/2 hours in September and a second period of like duration in late October. The use of two extended blocks of time early in the school year, when students are eager to get down to work, proved detrimental to morale, and possibly led to dubious responses because of that factor, in addition to producing test weariness from seven hours of testing in two days. Hence we decided to administer a number of the control measures, like IQ, later in the school year at a time when the youngsters enjoyed a diversion from the routines of the school day. This change, of course, had no bearing on the validity of the study since the measures administered at this time were no way influenced by the two months of instruction. Also, during this period of try-out, important revisions had to be made in the form used to collect data on students' prior achievement since guidance information, apparently, is maintained in many diverse ways and can be collected easily only when adequately planned for.

The second purpose of the pilot run was to check data processing procedures, such as the computer programs, which had to handle the large amount of data, and to explore whether each of the planned control measures independently influenced the final outcomes. The conclusion of the latter small investigation, which had to be carried out very quickly during the summer just preceding the final study, was the somewhat abbreviated design mentioned above.

1962-63 -- Instrument Construction

Also during the second year of the study, the remaining instrument had to be developed and/or put into final form. Thus, coincident with the pilot experiment, the unit tests on Periodic Table, on Water, Solutions, and Ionization, and on Principles of Reaction were administered, analyzed, and evaluated. The Instructional Process Report Form went through another trial and revision early in the academic year. The format that resulted from this revision was then used in the pilot study and proved satisfactory for the full study without further changes. The cost analysis procedures were explored with the aid of Professor M. Pullen, who served as consultant on this part, and a preliminary analysis of cost data was completed. Finally, the detailed schedule for 1963-64 was worked out after consultation with the officials of each of the schools. It incorporated the features which resulted from the pilot study experience, as well as the limitations and concerns required by the location of the study within operating high schools. The final effort of the year was the writing of a small, one-page description of the study to explain its nature to the youngsters who were to become its subjects, to their parents, and to their other teachers.

1963-64 -- Testing Sessions

The full study got under way almost on the first day of instruction in September 1963 with the Chemistry Achievement Pretest, a 15-minute test, administered by the collaborating teachers before any instruction had taken place. The other instruments in the pretest battery followed soon thereafter. A battery consisting of Science Reasoning and Understanding Test (50 min.), Parts I, II, and III of the Iowa Chemistry Aptitude Test (40 min.), and the Kuder Preference Record, scheduled for 60 minutes but for which time was allowed until the students completed the entire test, was administered by the study staff in a 3-1/2-hour testing period sometime during the week of September 9, 1963. Other control data were obtained during a second 3-1/2-hour testing session administered by the study staff during the week of October 28, 1963. This second set of tests consisted of the Holzinger-Crowder Unifactor Tests (90 min.) and the Critical Thinking Test (55 min.). The information form which recorded certain personal data as well as the students' past course achievement and the parents' education and occupations was inserted between these two tests (15 min.). The students were allowed excess time to complete the Critical Thinking Test. One control and two criterion measures were obtained in a third 3-1/2-hour testing session, late in the school year, at a time when all teachers had begun the customary pre-Regents Examination review. In the period between May 21 and 29, the study staff administered the Brown-Holtzman Survey of Study Habits and Attitudes (20 min.) and readministered the Science Reasoning and Understanding Test and the Kuder Preference Record.

All of the tests administered during the test sessions were scored by the study staff and results were furnished to the guidance officials of the individual schools as well as to the teachers. The final criterion instrument, the Regents Chemistry Examination of 1964, was scored by the individual teachers and the scores were obtained from them.

1963-64 -- Unit Tests

Each of the unit tests was administered by the individual teachers at the time when they had completed work on that particular unit. Since teachers spent various times on the materials covered on each of the unit tests, this led to a rather diverse schedule. However, since the staff attempted to measure outcomes of the kinds of courses which are typically taught by chemistry teachers, no restrictions were put on time, or any other factors related to the 15 teachers' individual teaching plans. Table III shows the wide variation in the time for completion of the relevant units. Indeed, at least one school inverted the order of presenting some of the units.

Teachers received keys to the unit tests and scored them for their own

TABLE III

SCHEDULE OF TEST ADMINISTRATIONS

SCHOOL	PRETEST	PERIODIC	WATER, SOL. and IONIZATION	HALOGENS	SULFUR-NIT.	PRINC. REACT.	ORGANIC
Akron	9/5/63	10/ 9/63	1/28/64	4/15/64	5/13/64	5/19/64	3/25/64
Alden	9/5/63	10/16/63	1/13/64	2/10/64	3/19/64	4/15/64	5/ 8/64
Amherst	9/5/63	10/24/63	1/16/64	3/ 5/64	4/21/64	5/26/64	5/13/64
Ben Franklin	9/10/63	10/ 9/63	12/17/63	2/11/64	3/20/64	3/11/64	5/18/64
Brighton	9/9/63	10/23/63	12/17/63	2/28/64	4/ 9/64	5/25/64	5/ 7/64
Canandaigua	9/6/63	10/17/63	2/14/64	3/20/64	4/10/64	3/19/64	5/20/64
Clarence	9/5/63	10/ 9/63	12/18/64	3/ 4/64	2/12/64	4/ 8/64	5/ 6/64
Cleveland Hill	9/6/63	10/11/63	12/ 5/63	1/16/64	3/17/64	5/20/64	4/21/64
Eden	9/5/63	9/27/63	11/ 1/63	1/13/64	2/26/64	5/22/64	5/ 1/64
Hamburg	9/5/63	9/25/63	12/ 4/63	2/11/64	3/24/64	4/13/64	5/ 1/64
Kenmore	9/9/63	11/ 1/63	1/22/64	2/14/64	3/18/64	5/20/64	5/ 1/64
Orchard Park	9/5/63	9/23/63	10/15/63	11/15/63	12/19/63	1/21/64	2/27/64
West Seneca	9/6/63	10/23/63	1/16/64	2/20/64	3/11/64	5/21/64	5/ 7/64
Williamsville	9/5/63	10/ 3/63	12/10/63	1/21/64	2/25/64	5/14/64	4/ 9/64

records. They were rescored by the study staff because for the purposes of the study, breakdown of items by item type was needed. Each of the teachers was also furnished with normative information on the unit tests from the accumulated data to translate their test scores into percentages. All teachers used the unit test grades as a significant part of the course grade so that motivation to do well on the tests was kept high. By and large, teachers "liked" the tests and found no difficulty in using them as major evaluation devices. Both students and teachers recognized the presence of questions that required cognitive competencies other than memorization for the tests through comments that referred to them as "reasoning" or "critical thinking" tests. That was, of course, one of the aims in their construction.

The test on Principles of Reaction was an exception to the general plan to administer the unit tests upon completion of each unit. Because of an earlier misunderstanding, or because changes had occurred in teaching plans between the planning year and 1963-64, it was discovered that the participating teachers typically do not teach a unit on principles of reaction but incorporate the contents of that unit into their teaching programs as the opportunity arises. Thus, it was impossible to find a block of content of some weeks duration after which that unit test would ideally fit, nor, as mentioned above, was it desirable to require changes in the teaching plans to establish such a block of content. Consequently, the participating teachers were directed to teach their normal courses and to administer the Principles of Reaction Test whenever they felt that they had adequately taught its contents. As can be seen from the test administration schedule which was actually followed, there was some homogeneity in the administration of that unit test also.

Teachers knew the content of each unit test from the tables of specification, and were sent the actual test booklets at least one week in advance of the day for which they required the test in their normal teaching program. They were asked not to review unusually for these unit tests--that is, not to "teach for the tests."

Instructional Process Report Form

Teachers were instructed to complete the Instructional Process Report Form from the first day of teaching until the last day of teaching for out-of-class contacts related to their chemistry courses as well as for their normal assigned classes. In order to reduce somewhat the total burden of reporting, a sampling procedure was established for class contacts by requesting a report for one section of Regents chemistry only in the conventionally organized schools, although all of the participating teachers' sections were part of the study. The large group - small group schools reported for the scheduled lecture, discussion, and laboratory meetings of one student only, which produced the equivalent of "one section." Since all but one school scheduled Regents chemistry for six or seven periods per week, the sampling procedure resulted in at least one and sometimes two reports for scheduled classes each day. With reports for the non-scheduled contacts, the average number of reports per teacher per day rose to about three.

Sample Characteristics

Tables IV and V list the schools and give some of the characteristics of the instructional groups which were the sample for this study during November 1963. (Some changes in enrollment occurred during the school year.) The conventional classes all met for five class periods per week and had one or

TABLE IV

CHARACTERISTICS OF CONVENTIONAL SCHOOLS
November 1963

SCHOOL	NUMBER OF CLASSES	ENROLLMENT EACH CLASS	TOTAL	CLASS MEETINGS PER WEEK	MINUTES CLASS PERIOD	LABS PER WEEK
Akron	3	28 16 12	56	5	46	2
Alden	3	19 24 19	62	5	52	2
Brighton	2	31 30	61	5	45	1
Cleveland Hill	2	19 15	34	5	47	1
	5	21 28 25 23 25	122	5	47	1
Eden	4	23 17 17 28	85	5	42	2
Hamburg	3	28 27 30	85	5	45	1
Orchard Park	3	26 28 26	80	5	46	2
West Seneca	3	27 20 29	76	5	45	2

TABLE V

CHARACTERISTICS OF LARGE GROUP-SMALL GROUP SCHOOLS
November 1963

	LARGE GROUP MEETINGS		SMALL GROUP MEETINGS			
	SIZE	MEETINGS PER WEEK	SIZES	MEETINGS PER WEEK	LAB MEETINGS PER WEEK	LENGTH OF PERIOD IN MINUTES
Amherst	153	4	15 14 16 14 15 15 11 16 17 20	1	1	48
Candaigua Academy	80	3	9 19 18 13 21	1	1	45
Clarence	139	2	32 27 15 9 23 21 12	2	2	50
Kenmore East Senior High	100	3	25 13 32 30	2	1	50
Ben Franklin Rochester, N.Y.	100	5	28* 24* 26* 22* (Laboratory Only)	-	2	47
Williamsville	78	2	17 21 19 21	3	1	49

*Lab only

two additional laboratory periods. Class length was typically 45-47 minutes, with one school each with 42 and 52 minute periods. Class sizes varied quite considerably, both between and within schools, from a low of 12 to a high of 31 pupils.

Variation was greater among the large group - small group schools. One school was exclusively large group, five periods per week, with small groups meeting only for laboratory. Among the remaining schools two met for two large group lectures each week, two for three large group meetings, and one for four. One or two small group meetings were typical although one school scheduled three such small groups each week. Two of the schools held two laboratory sessions each week, four only one. The size of the large group varied from 78 to 153 and the length of the lecture period from 45 to 50 minutes. The size of the small groups also varied considerably, from an atypical low of 7 to a high of 32. Thirty percent of the small groups contained 20 or more students which raises some questions concerning their "small group" character.

It should be noted that the teachers in the schools with apparently very low enrollments had other instructional assignments and that an additional teacher helped with some small group work in the two largest schools.

While the study sampled schools which had chosen one of the two organizational patterns by criteria external to the study and were thus separated by what might be termed a receptivity to innovation in the organization of instruction, the schools were not otherwise easily distinguishable. Both groups contained large and small schools; both contained schools in urban, suburban, small town, and rural settings. The schools appear similar in environmental and, presumably, school-climatic factors. They appear drawn from equally stratified populations.

Teacher factors likewise did not introduce any obvious skewness. All teachers had extensive experience in teaching the Regents chemistry course, and their students had established good records on past Regents Chemistry Examinations.

Finally, the students in the two groups were distinguishable to a statistically significant extent only on two of the eight measures used in the analysis of covariance (Table VI), and both the significant and the non-significant differences in mean scores favored one group in half the cases and the other in the other half. Hence, no systematic distinction between students could be identified. The two findings of significant differences are difficult to interpret. At least one is sufficiently large (0.01 level) to be unlikely as a result of chance. Nevertheless, even unlikely events happen at times, and this difference might represent such an unlikely event. Any effect it may have had on outcomes was corrected for by the statistical techniques used in the analysis of the data.

TABLE VI

COMPARISON OF MEAN SCORES ON EIGHT CONTROL VARIABLES

NAME OF TEST	LARGE GROUP PUPILS			CONVENTIONAL PUPILS			DIFF.	ST. ERROR OF DIFF.	LEVEL OF SIGNIF.
	N		SD	N		SD			
Scholastic Average	597	81.54	7.92	646	81.12	7.96	0.42	0.45	NS
Kuder, Science	655	41.57	15.34	674	41.84	15.70	0.27	0.85	NS
Chemistry Pretest	650	20.05	7.77	673	20.82	6.46	0.77	0.39	0.05
Science Reasoning Pretest	650	23.60	7.99	673	22.96	7.67	0.64	0.43	NS
Holzinger-Crowder: Verbal Scale	609	62.18	12.81	621	59.81	13.08	2.37	0.74	0.01
Spatial Scale	609	83.99	23.67	621	84.02	24.18	0.03	1.37	NS
Numerical Scale	609	84.59	17.73	621	84.67	17.21	0.08	1.00	NS
Reasoning Scale	609	67.00	12.69	621	65.86	13.23	1.14	0.74	NS

Chapter 4

PUPIL ACHIEVEMENT

Differences in the effectiveness of the two organizational patterns for achieving learnings of various sorts were evaluated by means of 15 different measures:

a. The Regents Examination, Parts I and II, and the Regents Examination total score;

b. Six units tests;

c. Items of all unit tests sorted according to four different intellectual behaviors required for their solution: recall, comprehension, application, and higher order intellectual behaviors;

d. Kuder Preference Record, Science Scale; and

e. Test of Science Reasoning and Understanding.

In order to take account, statistically, of initial differences among the students, analysis of covariance procedures were used, with eight independent measures as the control variables:

a. The Kuder Preference Record, Science Scale, taken as a pretest;

b. The Test on Science Reasoning and Understanding, taken as a pretest;

c. The Chemistry Achievement Pretest;

d. The Holzinger-Crowder Unifactor Tests: Verbal, Spatial, Numerical, and Reasoning; and

e. The Pupils' Scholastic Average for all academic subjects taken in 9th and later grades up to but not including courses taken during the year of the study, 1963-64.

These analyses determined differences in achievement under the two organizational plans without specific reference to the kind of learner involved.

They compare mean scores adjusted for differences between pupils on control measures.

Control measures make a maximum independent contribution the analysis if the correlation coefficients between them are small. Table VII gives the intercorrelations between the eight control variables. Except for the intercorrelations on the four scales of the Holzinger-Crowder test, the correlation coefficients were sufficiently low to justify their use as independent control measures.

Covariance procedures do not permit the assessment of differences in the effect of the organizational pattern on pupils with different competencies. To ascertain whether bright or slow students fared better under one plan or the other, regression analysis was used. This technique derives mathematical relationships between the measures that describe the students and the measures that determine their achievement. The coefficients of these mathematical relationships are then compared for significant differences, and if such differences exist, the nature of the effect can be investigated.

The sample of pupils to which these statistical procedures was applied was drastically reduced from the total number of pupils in the study since each pupil whose data were used in the final analysis had to have taken every test. Hence, pupils transferring out of the class, or into the class, or pupils who were ill or truant during any of the testing sessions were necessarily eliminated from the comparisons. Both the analysis of covariance and the regression analysis, therefore, was based on data with "healthy non-truants." While these students generally differed significantly in achievement measures from the total group who took the examination, we do not believe that this factor invalidates the results of the analysis. This particular problem is

TABLE VII

CORRELATION COEFFICIENTS BETWEEN EIGHT CONTROL VARIABLES

Large Group Pupils

	N	KUD	CHEM	SCI	H-C,V	H-C,S	H-C,N	H-C,R
1. Schol. Ave.	597	-05	33	41	41	14	22	27
2. Kuder, Sci. Pre.	655	-	14	36	-03	07	02	-01
3. Chem. Pretest	650		-	18	20	13	10	14
4. Sci. Reas. Pre.	650			-	31	13	10	33
5. H-C, Verbal	646				-	63	70	40
6. H-C, Spatial	646					-	64	46
7. H-C, Numerical	646						-	45
8. H-C, Reasoning	646							-

discussed in detail below.

Achievement -- The Regents Examination

Tables VIII a-c give the results of the analysis of covariance using the Regents Examination as the measure of course achievement. Part I of the Regents Examination is a 60-point objective test with items selected from all parts of the Regents Syllabus. Usually recall is heavily represented among these items; however, in recent years there has been a change in the examination toward items requiring other intellectual competencies. The report of pupils and teachers regarding the 1964 Regents Examination was that it was a "harder" "thinking-type" examination; that is, it probably contained a noticeable fraction of comprehension, application, analysis, synthesis, or evaluation-type items although the construction of the test does not follow a design which would assure a desired distribution of different kinds of items. Part II of the Regents Examination consists of fewer, more extensive questions with a choice among the total number of items. The scoring of Part II is objective also.

For boys, there are a small, though apparently non-chance difference in the scores on Part I of the Regents Examination and larger differences on Part II and on the total Regents score. Both are in favor of the group that received instruction under the large group pattern. For girls, the differences are not large enough to be statistically significant; however, they are in the same direction as for boys. The differences are maintained for all students compared regardless of sex and are large enough to be statistically significant. Hence, based on our data and within the limitations of our sample and design, a decision to instruct under the large group - small group organizational

TABLE VIII-a

ANALYSIS OF COVARIANCE WITH PART ONE OF THE JUNE, 1964, REGENTS EXAMINATION IN CHEMISTRY AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	44.62	8.07	43.91	6.966	.01
Conventional Pupils	457	42.20	9.93	42.74		
Large Group Females	172	44.10	8.32	42.96	1.744	NS
Conventional Females	200	41.13	9.23	42.11		
Large Group Males	172	45.14	7.81	44.80	6.516	.05
Conventional Males	257	43.03	10.39	43.25		
All Female Pupils	372	42.51	8.93	42.05	18.121	.01
All Male Pupils	429	43.87	9.49	44.27		
TOTAL SAMPLE	801	43.24	9.25			

TABLE VIII-b

ANALYSIS OF COVARIANCE WITH PART TWO OF THE JUNE, 1964, REGENTS EXAMINATION IN CHEMISTRY AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	31.43	6.78	30.94	7.601	.01
Conventional Pupils	457	29.43	8.37	29.81		
Large Group Females	172	31.24	7.08	30.47	0.236	NS
Conventional Females	200	29.50	8.21	30.16		
Large Group Males	172	31.62	6.48	31.42	12.588	.01
Conventional Males	257	29.38	8.51	29.51		
All Female Pupils	372	30.31	7.75	29.97	1.493	NS
All Male Pupils	429	30.28	7.83	30.47		
TOTAL SAMPLE	801	30.29	7.78			

TABLE VIII-c

ANALYSIS OF COVARIANCE WITH TOTAL SCORES ON THE JUNE, 1964, REGENTS EXAMINATION IN CHEMISTRY AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	76.01	14.00	74.80	8.513	.01
Conventional Pupils	457	71.62	17.65	72.53		
Large Group Females	172	75.27	14.51	73.36	0.971	NS
Conventional Females	200	70.58	16.70	72.22		
Large Group Males	172	76.74	13.48	76.21	10.737	.01
Conventional Males	257	72.43	18.34	72.79		
All Female Pupils	372	72.75	15.88	71.94	10.133	.01
All Male Pupils	429	74.16	16.68	74.86		
TOTAL SAMPLE	801	73.50	16.32			

pattern is educationally defensible if higher achievement in the Regents Examination is the desired criterion.

The data also confirm the generally accepted notion that boys perform better than girls in chemistry. Boys' achievement was substantially and significantly higher on Part I of the Regents Examination, and this difference carried over to the Total Score of the Regents Examination. On Part II, the sex difference was not statistically significant. This finding requires further comment later since its implications are contrary to others found in this study.

Achievement -- The Unit Tests

Tables IX a-f give the results of the comparisons of achievement scores on the six unit tests constructed for this study. These tests are particularly appropriate as measuring instruments since their items followed the tables of specifications of the teachers who participated in the study. They reflected the teaching sequence of the teachers in greater detail than the Regents Chemistry Syllabus. On every test and for every comparison between large group and conventially-taught students, there is a significant difference in the adjusted means of the achievement scores in favor of the large group students. The differences approach one half standard deviation, leading to F-ratios of such magnitude that a treatment effect seems beyond question. Again within the limitations of the present study, the achievement on each of the six unit tests shows that a decision to instruct under the large group pattern is educationally defensible.

It is interesting that as with the results on the Regents Chemistry Examination, only some of the unit tests show significant differences between males and females. On one unit test the difference is in favor of boys at the

TABLE IX-a

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING THE PERIODIC TABLE OF ELEMENTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	34.53	6.92	34.00	49.525	.01
Conventional Pupils	457	30.90	7.87	31.30		
Large Group Females	172	34.90	6.63	34.09	29.393	.01
Conventional Females	200	30.44	7.13	31.14		
Large Group Males	172	34.17	7.20	33.87	19.880	.01
Conventional Males	257	31.25	8.40	31.45		
All Female Pupils	372	32.50	7.24	32.26	0.668	NS
All Male Pupils	429	32.42	8.06	32.64		
TOTAL SAMPLE	801	32.46	7.69			

TABLE IX-b

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING WATER, SOLUTIONS AND IONIZATION AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	33.21	6.98	32.45	84.203	.01
Conventional Pupils	457	28.75	7.94	29.17		
Large Group Females	172	33.02	6.71	33.02	52.141	.01
Conventional Females	200	27.43	7.36	27.43		
Large Group Males	172	33.29	7.25	33.12	36.095	.01
Conventional Males	257	29.79	8.23	29.97		
All Female Pupils	372	30.02	7.59	29.94	8.412	.01
All Male Pupils	429	31.23	8.04	31.30		
TOTAL SAMPLE	801	30.67	7.85			

TABLE IX-c

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING THE HALOGEN ELEMENTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	35.40	7.02	34.82	109.438	.01
Conventional Pupils	457	30.42	8.13	30.86		
Large Group Females	172	35.38	7.09	34.54	43.535	.01
Conventional Females	200	30.12	7.51	30.85		
Large Group Males	172	35.42	6.97	35.11	67.558	.01
Conventional Males	257	30.65	8.58	30.86		
All Female Pupils	372	32.55	7.77	32.12	3.037	NS
All Male Pupils	429	32.57	8.30	32.95		
TOTAL SAMPLE	801	32.56	8.05			

TABLE IX-d

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING THE SULFUR-NITROGEN FAMILIES AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	35.31	7.67	34.79	46.276	.01
Conventional Pupils	457	31.19	8.56	31.58		
Large Group Females	172	35.44	7.78	34.62	28.565	.01
Conventional Females	200	30.22	8.08	30.93		
Large Group Males	172	35.19	7.57	34.93	19.349	.01
Conventional Males	257	31.93	8.85	32.10		
All Female Pupils	372	32.63	8.35	32.23	5.664	.05
All Male Pupils	429	33.24	8.50	33.59		
TOTAL SAMPLE	801	32.96	8.43			

TABLE IX-e

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING THE PRINCIPLES OF CHEMICAL REACTIONS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	36.21	8.00	35.56	70.417	.01
Conventional Pupils	457	31.12	9.64	31.61		
Large Group Females	172	36.09	8.33	35.06	25.218	.01
Conventional Females	200	30.59	9.26	31.48		
Large Group Males	172	36.34	7.67	36.06	48.136	.01
Conventional Males	257	31.53	9.93	31.71		
All Female Pupils	372	33.13	9.25	32.63	4.719	.05
All Male Pupils	429	33.45	9.38	33.39		
TOTAL SAMPLE	801	33.31	9.31			

TABLE IX-f

ANALYSIS OF COVARIANCE WITH SCORES ON THE UNIT TEST COVERING ORGANIC CHEMISTRY AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	35.26	7.04	34.72	43.447	.01
Conventional Pupils	457	31.61	8.23	32.01		
Large Group Females	172	35.49	6.90	34.73	17.325	.01
Conventional Females	200	31.58	7.53	32.22		
Large Group Males	172	35.03	7.19	34.77	27.084	.01
Conventional Males	257	31.63	8.74	31.81		
All Female Pupils	372	33.38	7.50	32.92	0.931	NS
All Male Pupils	429	33.00	8.32	33.40		
TOTAL SAMPLE	801	33.18	7.94			

0.01 level of significance; two tests produced significant differences at the 0.05 level; and three tests show no statistically significant differences between the sexes

Achievement -- Intellectual Competencies

When the items of all six unit tests are grouped according to the intellectual competencies required for their solution, the same results are obtained as when the unit test items are grouped by topic. As Tables X a-d show, the differences are very large and in favor of large group instructed pupils on all four types of items--those requiring recall, comprehension, application, and higher competencies--whether comparisons involve males only, females only, or both sexes combined. As in the comparison with the unit test items grouped according to topic, the differences approach one half standard deviation and strongly suggest the existence of a treatment effect. Again within the limitations of the current study, a decision to instruct under the large group pattern is educationally defensible whether the objective is the learning of recall, comprehension, application, or analysis, synthesis and evaluation.

It is of extreme interest that when the comparisons are made by sex alone regardless of the organization of the instruction, there are no significant differences on recall and comprehension items but significant differences in favor of males on application and higher competence items. These data might be interpreted as a male superiority on those intellectual competencies specifically needed for the problem solving tasks that are integral to chemistry and which are rated "higher" in the hierarchy of intellectual behaviors. We feel that such an inference would be rash. We accept as incontrovertible the data that show superior male achievement; however, the observed differences

TABLE X-a

ANALYSIS OF COVARIANCE WITH TOTAL SCORES ON TYPE ONE ITEMS ON SIX UNIT TESTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	70.97	10.82	69.95	99.302	.01
Conventional Pupils	457	62.90	14.42	63.67		
Large Group Females	172	71.30	11.00	69.84	48.261	.01
Conventional Females	200	62.13	13.11	63.39		
Large Group Males	172	70.63	10.67	70.15	53.722	.01
Conventional Males	257	63.50	15.36	63.82		
All Female Pupils	372	66.37	13.00	65.68	2.606	NS
All Male Pupils	429	66.36	14.10	66.96		
TOTAL SAMPLE	801	66.37	13.59			

TABLE X-b

ANALYSIS OF COVARIANCE WITH TOTAL SCORES ON TYPE TWO ITEMS IN SIX UNIT TESTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	70.71	12.59	69.60	70.674	.01
Conventional Pupils	457	63.27	15.02	64.10		
Large Group Females	172	70.66	12.64	68.87	27.543	.01
Conventional Females	200	62.40	13.88	63.94		
Large Group Males	172	70.66	12.58	70.24	42.340	.01
Conventional Males	257	63.95	15.85	64.29		
All Female Pupils	372	66.22	13.93	65.71	3.093	NS
All Male Pupils	429	66.68	14.99	67.12		
TOTAL SAMPLE	801	66.47	14.50			

TABLE X-c

ANALYSIS OF COVARIANCE WITH TOTAL SCORES ON TYPE THREE ITEMS ON SIX UNIT TESTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	40.92	9.74	40.13	102.138	.01
Conventional Pupils	457	34.14	11.49	34.74		
Large Group Females	172	40.95	10.01	39.75	57.480	.01
Conventional Females	200	32.77	10.50	33.81		
Large Group Males	172	40.88	9.50	40.45	47.133	.01
Conventional Males	257	35.21	12.13	35.50		
All Female Pupils	372	36.55	11.05	36.01	8.624	.01
All Male Pupils	429	37.48	11.48	37.96		
TOTAL SAMPLE	801	37.05	11.28			

TABLE X-d

ANALYSIS OF COVARIANCE WITH TOTAL SCORES ON TYPE FOUR ITEMS ON SIX UNIT TESTS AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	27.33	5.59	26.87	89.465	.01
Conventional Pupils	457	23.67	6.35	24.02		
Large Group Females	172	27.40	5.72	26.73	44.200	.01
Conventional Females	200	23.08	5.82	23.66		
Large Group Males	172	27.27	5.49	27.02	49.017	.01
Conventional Males	257	24.12	6.70	24.29		
All Female Pupils	372	25.08	6.15	24.69	7.566	.01
All Male Pupils	429	25.38	6.42	25.72		
TOTAL SAMPLE	801	25.24	6.30			

are probably caused by factors other than a male superiority on complex intellectual behaviors with no difference between the sexes on simpler intellectual tasks. We reach this conclusion for at least two reasons. The data on the Regents Examination were exactly converse to those obtained here. On the Regents Examination, boys excelled on Part I, the more rote part, and did not achieve significantly higher scores than girls on Part II, the problem solving part. Moreover, other data that arose from this study strongly imply that the separate intellectual competencies that emerged from our categorizing scheme are not psychologically different from each other. As we discuss below, their differences seem to reside in a priori logic, not in empirical fact.

Achievement -- Science Reasoning

In order to determine whether any of the gains achieved in the chemistry course could be considered transferrable rather than specific for the syllabus content, the Test of Science Reasoning and Understanding was included among the measures of achievement. It could be assumed that any differences on this instrument indicate changes, as a result of instruction, in the ability to apply intellectual behaviors in areas relevant and related to the chemistry course, a minimum psychological distance from the content of the course, yet not specifically related to chemistry. Unfortunately, the Test of Science Reasoning and Understanding is only vaguely defined and is no longer generally available. Generalized science reasoning is probably a fictitious intellectual category (Guilford), and there are few good measurement instruments that remain from the time when its existence was postulated without question. Thus in the time since the initial planning of this study, evidence has come to light which throws some doubts on the interpretation of data on this test.

Table XI gives the results of the covariance analysis with the Test of Science Reasoning and Understanding. No significant differences can be found on any of the comparisons. Neither method of organizing the instruction differed in a statistically significant way from the other in inducing generalized reasoning with respect to science problems. This does not imply that the year of instruction had no effect. A comparison between mean scores obtained in September and in May (Table XII) clearly shows substantial gains for both groups. But the groups were not affected differentially when corrections were made for initial differences.

The data on the Test of Science Reasoning and Understanding require additional consideration. The analysis of covariance procedure used the same measure to correct outcomes for initial differences among the pupils. While the correlation coefficients between pretest and posttest were small, 0.30 and 0.29 for conventional and large group pupils respectively, the analytical technique is extremely sensitive and requires very large treatment effects before statistical significance would appear. Perhaps the analytical procedure obscures a real effect. This speculation receives some support from the raw data used without covariance adjustment which, however, suffer from the defect that the students who took the test in September differed somewhat from the students who took the test in May because of in-migration, out-migration, and absences on the two testing days. If the uncorrected data for all pupils are representative of the actual results of instruction under the two organizational patterns, then science reasoning is the only measure in this study in which conventionally taught pupils achieved significantly higher scores at the end of instruction than large group - small group taught pupils.

The various ways in which our data can be interpreted suggest that the

TABLE XI

ANALYSIS OF COVARIANCE WITH SCIENCE REASONING AND UNDERSTANDING POSTTEST AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	344	29.68	5.67	29.16	3.406	NS
Conventional Pupils	457	28.21	5.96	28.61		
Large Group Females	172	29.49	5.58	28.80	2.325	NS
Conventional Females	200	27.56	4.89	28.15		
Large Group Males	172	29.88	5.78	29.59	2.741	NS
Conventional Males	257	28.72	6.64	28.91		
All Female Pupils	372	28.45	5.30	29.29	2.741	NS
All Male Pupils	429	29.18	6.33	28.81		
TOTAL SAMPLE	801	28.84	5.88			

TABLE XII

COMPARISON OF SCORES ON TEST OF SCIENCE REASONING AND UNDERSTANDING TAKEN BEFORE AND AFTER THE YEAR OF INSTRUCTION, ALL STUDENTS

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	DIFFERENCE	ST. ERROR OF DIFF.	LEVEL OF SIGNIFICANCE
Large Group - Sept.	650	23.60	7.99	0.64	0.43	NS
Conventional - Sept.	673	22.96	7.67			
Large Group - May	565	26.65	9.46	1.27	0.48	0.01
Conventional - May	598	27.82	6.34			
Large Group - Sept.	650	23.60	7.99	3.05		
Large Group - May	565	26.65	9.46			
Conventional - Sept.	673	22.96	7.67	4.86		
Conventional - May	598	27.82	6.34			

achievement of non-specific science-related intellectual behavior, as represented by the Test of Science Reasoning and Understanding, is an area in which additional research would be fruitful. The preliminary conclusions, in view of the design of our study, is that neither group was affected differently from the other in the competencies measured by the Test of Science Reasoning and Understanding.

Achievement -- Science Interest

Motivation plays an important role in the advantages cited by proponents of the large group organization in the high school. Unfortunately, it is one of the most difficult things to measure. The best measure would be performance in later years--college preparation for and actual persistence into a science-related occupation--but data could not be attained for several years. (We are prepared for such a follow-up study in future years.) As an approximate indication of motivation toward science we chose the Kuder Preference Record, Science Scale. It was administered both before the study, to differentiate pupils, and after the study, to assess changes in outcome. As we indicated earlier, the two groups appeared undifferentiated on this measure before the study began.

The treatment apparently had little or no effect on science interest. The correlation between scores measured in September and in May was near 0.65 for all groups and sub-groups, showing a significant influence of initial interest on the interest score at the end of instruction. Moreover, there were only small gains on the science interest scores between pretest and posttest (Table XIII) for the uncorrected samples. The result persisted in the covariance analysis (Table XIV).

TABLE XIII

COMPARISON OF SCORES ON KUDER PREFERENCE RECORD, SCIENCE SCALE, TAKEN BEFORE AND AFTER THE YEAR OF INSTRUCTION, ALL STUDENTS

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	DIFFERENCE	ST. ERROR OF DIFF.	LEVEL OF SIGNIFICANCE
Large Group - Sept.	655	41.57	15.34	0.27	.85	NS
Conventional - Sept.	674	41.84	15.70			
Large Group - May	514	42.46	12.43	0.56	.79	NS
Conventional - May	591	41.90	13.79			
Large Group - Sept.	655	41.57	15.34	0.89		
Large Group - May	514	42.46	12.43			
Conventional - Sept.	674	41.84	15.70	0.06		
Conventional - May	591	41.90	13.79			

TABLE XIV

ANALYSIS OF COVARIANCE WITH KUDER PREFERENCE RECORD POSTTEST AS THE DEPENDENT VARIABLE AND SCHOLASTIC AVERAGE, KUDER SCIENCE, CHEMISTRY PRETEST, SCIENCE REASONING PRETEST AND HOLZINGER-CROWDER VERBAL, SPATIAL, NUMERICAL AND REASONING SCORES AS COVARIABLES

GROUP IDENTITY	NUMBER	MEAN	STANDARD DEVIATION	ADJUSTED MEAN	F-RATIO	LEVEL OF SIGNIFICANCE
Large Group Pupils	354	43.03	12.37	42.97	3.345	NS
Conventional Pupils	466	42.00	13.70	42.04		
Large Group Females	175	37.83	11.17	36.88	2.861	NS
Conventional Females	205	34.76	11.87	35.57		
Large Group Males	179	48.10	11.36	48.33	1.429	NS
Conventional Males	261	47.69	12.29	47.53		
All Female Pupils	380	36.17	11.64	41.07	18.061	.01
All Male Pupils	440	47.85	11.91	43.62		
TOTAL SAMPLE	820	42.44	13.14			

One difference, however, stands out from the covariance analysis. There is a significant difference between boys and girls in science interest at the end of the year despite the fact that their scores were adjusted for initial differences on the same measure. During the year they studied chemistry, whether by conventional or large group-small group instruction, girls became less interested in science while boys became more interested, at least as far as this interest is measured by the Kuder Preference Record. This finding has important implications for occupational guidance and considerations of manpower need and supply.

The Science Scale of the Kuder Preference Record shows no differences between conventional and large group taught pupils, no gains of interest when sexes are not separated, but a gain for boys and a loss for girls of statistical significance between the September and May.

Student Differences

In the comparisons we have discussed up to this point, we used data without regard for pupil differences except for sex. To determine whether two organizational patterns affect different sub-samples of students differently, we used regression analysis. We were concerned with whether one or the other kind of instruction favored students who scored high or low on any or all of the eight control variables.

As mentioned in the beginning of this chapter, regression analysis establishes mathematical equations that relate the control variables to achievement variables. The terms of the resulting equations for the two treatment groups are then compared and tested for statistically significant differences. With 8 predictor variables and 15 criterion variables, 120 simple regression equations and more than 1000 multiple regression equations result. Since the computer program required a score for every subject on each of the 23 measures, this analysis also was limited to the somewhat restricted sample of students who were neither ill nor truant during the year.

The number of regression coefficients produced by the analysis is too large to be reproduced in this report. All simple and all possible multiple regression equations were computed and their coefficients compared for the large group-small group and conventional students. Among the several thousand comparisons, only a handful gave differences that appeared statistically significant at the 0.05 per cent level. Since chance alone produces differences of that magnitude in one trial of 20, and since the actual data produced fewer than that fraction of apparently different regression lines, we attach no importance to the differences which appeared significant and ascribe them entirely to chance. The fact that no pattern could be detected among the few results that seemed significantly different lends further support to that inference. We could not identify any subgroup of students, with high IQ or lower IQ, with higher achievement records or lower achievement records, with better chemistry preparation or poorer chemistry preparation, or with greater or lesser science interest, for whom achievement results differed from the results obtained with the entire sample. (The only exception, as noted previously are male or female subgroups.) Hence, we conclude that there are no subgroups of students who are affected differently from the total sample by either organizational pattern for instruction.

Nevertheless, among the wealth of data accumulated in the course of the study, there are a number of indications concerning different effects which require further attention. In the first place, regardless of organizational pattern, the "healthy non-truants," those pupils for whom we have data on every test administered, are not a random sample of the total students in the study. Tables XV a, b show that for the large group taught students on every measure except the Kuder Science Scale, and for the conventially taught students on about two-thirds of the measures, the restricted, covariance analysis sample differed significantly from all the students who took the same tests. It suggests that students who are more diligent in their attendance achieve more highly, or that students who are migrant or ill are penalized in their achievement. That finding is not surprising, and since covariance analysis adjusts the data to offset problems arising from this source, it does not invalidate the conclusions we have drawn regarding achievement differences among the two kinds of teaching. But these data led us to examine the results obtained on each test for all students and to compare the data to see whether we could get any additional insights into just exactly what was happening to different students under the two organizational patterns. The fact that the initial samples appeared indistinguishable, with each group excelling on half the control measures and significantly higher on one, gave some justification for this unplanned analysis.

The effects found with the analysis of covariance on the restricted sample carry over to the total sample (Table XVI). The large group - small group students surpassed conventionally taught students on 13 of 15 comparison measures though the differences were statistically significant on only 9 of them. Differences were significant at the 0.01 level on three unit tests, on Type III items, and on Part I of the Regents Examination. Significance dropped to the 0.05 level for another unit test, Type IV items, and Part II and total score of the Regents Examination. On the other hand, the conventional students

TABLE XV-a

COMPARISON OF SAMPLE SELECTED FOR COVARIANCE ANALYSIS WITH TOTAL STUDENTS - LARGE GROUP INSTRUCTION

	TOTAL SAMPLE			COV. SAMPLE				
MEASURE	NO.	MEAN	S.D.	MEAN	S.D.	DIFF.	ST. ERROR OF DIFF.	LEVEL OF SIGNIF.
UNIT TESTS:								
Per. Table	629	30.80	9.79	34.53	6.92	3.73	.37	0.01
Water, Soln., &Ion.	629	28.36	12.56	33.21	6.98	4.85	.38	0.01
Halogens	628	28.69	14.62	35.40	7.02	6.71	.38	0.01
Prin. Reactn.	526	35.02	9.37	36.21	8.00	1.19	.43	0.01
Sulfur-Nitrogen	629	27.58	15.35	35.31	7.67	7.73	.41	0.01
Organic	527	33.87	8.74	35.26	7.04	1.39	.38	0.01
ITEMS:								
Type I	633	58.25	23.24	70.97	10.82	12.72	.58	0.01
Type II	633	58.00	24.26	70.71	12.59	12.71	.68	0.01
Type III	632	33.51	14.97	40.92	9.72	7.41	.52	0.01
Type IV	632	22.50	9.43	27.33	5.59	4.83	.30	0.01
REGENTS EXAM:								
Part I	541	43.39	8.72	44.62	8.07	1.23	.44	0.01
Part II	540	30.16	7.67	31.43	6.78	1.27	.37	0.01
Total	539	73.47	15.16	76.01	14.00	2.54	.75	0.01
SCI. REAS. POST	565	26.65	9.46	29.68	5.67	3.03	.31	0.01
KUDER, SCI., POST	514	42.46	12.43	43.03	12.37	0.57	.67	NS

(N for Covariance Sample is 344)

TABLE XV-b

COMPARISON OF SAMPLE SELECTED FOR COVARIANCE ANALYSIS WITH TOTAL STUDENTS - CONVENTIONAL INSTRUCTION

MEASURE	TOTAL SAMPLE NO.	TOTAL SAMPLE MEAN	TOTAL SAMPLE S.D.	COV. SAMPLE MEAN	COV. SAMPLE S.D.	DIFF.	ST. ERROR OF DIFF.	LEVEL OF SIGNIF.
UNIT TESTS:								
Per. Table	649	29.22	9.26	30.90	7.87	1.68	.37	0.01
Water,Soln.,&Ion.	649	26.94	9.38	28.75	7.94	1.81	.37	0.01
Halogens	649	27.99	10.40	30.42	8.13	2.43	.38	0.01
Prin. Reactn.	598	30.62	9.72	31.12	9.64	0.50	.45	NS
Sulfur-Nitrogen	649	28.04	11.68	31.19	8.56	3.15	.40	0.01
Organic	601	30.56	9.22	31.61	8.23	1.05	.38	0.01
ITEMS:								
Type I	653	57.30	18.72	62.90	14.42	5.60	.67	0.01
Type II	653	57.65	19.52	63.27	15.02	5.62	.70	0.01
Type III	653	31.07	12.88	34.14	11.49	3.07	.54	0.01
Type IV	653	21.65	7.67	23.67	6.35	2.02	.30	0.01
REGENTS EXAM:								
Part I	541	42.01	9.70	42.20	9.93	0.19	.46	NS
Part II	607	29.19	8.51	29.43	8.37	0.24	.39	NS
Total	606	71.23	17.43	71.62	17.65	0.39	.83	NS
SCI. REAS. POST	598	27.82	6.34	28.21	5.96	0.39	.28	NS
KUDER,SCI.,POST	591	41.90	13.79	42.00	13.70	0.10	.64	NS

(N for Covariance Sample is 457)

TABLE XVI

COMPARISON OF UNCORRECTED SCORES OF LARGE GROUP STUDENTS WITH CONVENTIONAL STUDENTS

MEASURE	CONV. STUDENTS NO.	CONV. STUDENTS MEAN	CONV. STUDENTS S.D.	LARGE GP. STUDENTS NO.	LARGE GP. STUDENTS MEAN	LARGE GP. STUDENTS S.D.	DIFF.	ST. ERROR OF DIFF.	LEVEL OF SIGNIF.
UNIT TESTS:									
Per. Table	649	29.22	9.26	629	30.80	9.79	1.58	.53	0.01
Water, Soln.,&Ion.	649	26.94	9.38	629	28.36	12.56	1.42	.62	0.05
Halogens	649	27.99	10.40	628	28.69	14.62	0.70	.88	NS
Prin. Reactn.	598	30.62	9.72	526	35.02	9.37	4.40	.57	0.01
Sulfur-Nitrogen	649	28.04	11.68	629	27.58	15.35	0.46	.76	NS
Organic	601	30.56	9.22	527	33.87	8.74	3.31	.54	0.01
ITEMS:									
Type I	653	57.30	18.72	633	58.25	23.24	0.95	1.18	NS
Type II	653	57.65	19.52	633	58.00	24.26	0.35	1.23	NS
Type III	653	31.07	12.88	632	33.51	14.97	2.44	.78	0.01
Type IV	653	21.65	7.67	632	22.50	9.43	0.85	.48	0.05
REGENTS EXAM:									
Part I	606	42.01	9.70	541	43.39	8.72	1.38	.54	0.01
Part II	607	29.19	8.51	540	30.16	7.67	0.97	.48	0.05
Total	606	71.23	17.43	539	73.47	15.16	2.24	.96	0.05
SCI. REAS. POST	598	27.82	6.34	565	26.65	9.46	1.27	.47	0.01
KUDER, SCI., POST	591	41.90	13.79	514	42.46	12.43	0.56	.79	NS

scored higher on the Test of Science Reasoning and Understanding at the 0.01 level of significance, and non-significantly on the Sulfur-Nitrogen test.

The uncorrected data generally support our initial conclusions regarding higher achievement in the large group - small group schools. But it is difficult to see from the data just what is happening. The restricted large group sample was somewhat more select than the restricted conventional sample and it also was relatively smaller. Only 64% of the large group pupils who took the Regents Examination were in the restricted sample in contrast to 75% of the conventional pupils. Attendance was apparently better in conventional classes. Perhaps more importantly, persistence in the course was also greater in the conventional classrooms (Table XVII). The net decrease in enrollment from September to June, as determined by the number of students who took our tests, was 10% for conventional classes whereas the net decrease was 17% for the large group - small group classes. While these figures need further study since in-migration, out-migration, and illness affect them and since counselors might hesitate to place transfer students into large groups in mid-term, we feel that a significant residual effect probably remains which simply drives a larger fraction of students from large group taught chemistry classes. The relative losses between October, by which time classes typically have stabilized, and June are 2.3% and 10.8% respectively. The data signify, we believe, that students withdraw from large group - small group classes in significantly greater numbers than from conventional classes.

The last few paragraphs have thrown some light on a number of factors which are of great educational significance and which deserve to be investigated more fully. Quite obviously, even the exhaustive analysis of achievement measures which was undertaken in this study does not fully answer all questions

TABLE XVII

NUMBER OF STUDENTS WHO TOOK TESTS IN SEPTEMBER, OCTOBER AND JUNE

DATE AND TEST	CONVENTIONAL			LARGE GROUP		
	N	LOSS	%LOSS	N	LOSS	%LOSS
First week of school (Chem. Pretest)	673			650		
Late October (Holzinger-Crowder)	621	52	7.7	609	41	6.3
Late June (Regents Examination)	606	67	10.0	539	111	17.1

SIGNIFICANCE OF DIFFERENCES BETWEEN LOSS RATES UNDER THE TWO ORGANIZATIONAL PATTERNS

DATES COMPARED	CHI-SQUARE	LEVEL OF SIGNIFICANCE
First week of school and late October	0.8	NS
Late June and first week of school	15.0	0.01
Late June and late October	73.5	0.01

one must raise with regard to large group - small group instruction. Other values enter the picture. We now know that the holding power of the large group classes is poorer. We have some vague intimations that it may work preferentially against poorer students although our data are insufficient to clear up this point. But we do not know how poor these drop-outs are. Are they so poor that they would hardly have profited from persistence, or were they poor but generally successful students? Further data would help answer some of these vital questions.

Classification of Intellectual Behaviors

Several times already we have cast doubts on inferences with regard to different intellectual competencies although in several places in our initial research design, we assumed that such competencies exist and can be measured. Subsequent experimental results, however, have caused a change in our assumptions.

In the first place, there were contradictory results from the Regents Examination, on which males surpassed females only on the memory part, and from the various levels of intellectual complexity on the unit tests, on which males surpassed females on application and higher order competencies but not on memory. Secondly, there was the Test on Science Reasoning and Understanding which did not corroborate the significant differences in favor of large group - small group students found on supposedly comparable higher order competency items. Finally, there is the recent work of Guilford on the structure of creative intellectual processes which shows them to be quite complex, more complex than the theoretical structure developed by Bloom and used as the basis for our differentiations. But, we have direct evidence that our adaptation of Bloom's categories has no empirical basis. Intercorrelation coefficients between the items of Types I, II, III, and IV, for both conventional and large group pupils, are not less than .80 (Table XVIII). Hence, the items reach nearly the limit of reliability of any set of related items and differentiations among them hardly exist. Correlations between the items of various types and other measures, likewise, do not distinguish among them. The correlation coefficients are almost identical for items of Type I, II, III, and IV and each of the control measures used in this study (Table XIX). Indeed, it seems particularly significant that the correlation between the Holzinger-Crowder Reasoning Score and our item Type IV is non-significantly larger than the correlation between that score and our item Type I.

There can be no question that our items do not empirically differentiate the presumed intellectual behavior categories adapted from Bloom's taxonomy. There are several different interpretations possible, however. It may be that our items were incompetently constructed for the purposes at hand. This is possible, though a jury of 14 teachers agreed with little difficulty to distinctions between recall and the other three item types (if not as easily distinctions among the other item types). Yet, we find no empirical distinction even between items of Type I and the other types. Another explanation suggests that there is a general factor operative among all items in addition to the distinguishing mental competencies of the more complex types. Recall is admittedly required for all items and it may affect them to an overriding extent, or some other general factor may affect them to an overriding extent. This is quite likely. However, it makes the categories operationally indistinguishable and therefore theoretically useless. Whatever the reason for the high intercorrelations, we feel that our data demonstrate very clearly that the categories we used are

TABLE XVIII

INTERCORRELATION MATRIX BETWEEN RECALL, COMPREHENSION, APPLICATION, AND HIGHER ORDER COMPETENCY ITEMS

	II	III	IV
Item Type I Recall	.87	.84	.81
Item Type II Comprehension	-	.84	.80
Item Type III Application		-	.82
Item Type IV Higher Order Competencies			-

TABLE XIX

CORRELATION COEFFICIENTS BETWEEN ITEM TYPES AND OTHER MEASURES

	ITEM TYPE LARGE GROUP PUPILS				ITEM TYPE CONVENTIONAL PUPILS			
OTHER MEASURES	I	II	III	IV	I	II	III	IV
Chemistry Regents	.64	.69	.70	.60	.75	.77	.74	.71
Scholastic Average	.49	.51	.55	.50	.54	.56	.55	.56
Kuder Science, Pretest	.09	.10	.10	.09	.25	.23	.22	.24
Chemistry Pretest	.37	.35	.39	.40	.52	.55	.53	.52
Science Reasoning Pretest	.21	.24	.28	.26	.37	.36	.39	.40
Holzinger-Crowder								
Verbal	.27	.27	.31	.28	.47	.48	.43	.45
Spatial	.18	.22	.22	.19	.32	.36	.33	.32
Numerical	.29	.29	.32	.27	.39	.41	.42	.36
Reasoning	.22	.26	.31	.27	.37	.42	.44	.42

N for Large Group -- 609 or more

N for Conventional -- 621 or more

Type I - Recall

Type II - Comprehension

Type III - Application

Type IV - Higher Intellectual Behaviors

empirically indistinguishable and that their use in learning and curriculum theory places a psychologically non-existent phantom into the discussions. We would suggest that there are, no doubt, intellectual competencies of various orders but that the words recall, comprehension, application, analysis, synthesis, and evaluation are inadequate to describe them. This too is an area in which much valuable research remains to be done.

Chapter 5

PREDICTING ACHIEVEMENT

The equations relating pretest scores with achievement scores showed no difference between large group - small group pupils and conventional pupils. But, the correlation coefficients were high in many cases. This suggested the possibility of making a fairly reliable prediction of Regents Examination scores early in the school year from one or two test scores. Such prediction, of course, assumes that the difficulty of the Chemistry Regents Examination will remain stable. Since the 1964 examination was judged to be relatively difficult, such an assumption is probably justified although if the difficulty level should drop, prediction based on the 1963-64 data would tend to underestimate achievement. That could be considered an advantage since a final score higher than predicted may be considered more desirable in practice than the reverse. If the Regents Examination should become more difficult, on the other hand, the use of the data of this chapter would probably be unjustified.

Anyone who attempts to make a prediction about individual behavior is dealing with tenuous probabilities, and this is perhaps more true for a prediction of a score on a single examination that for a prediction of overall success in high school. Nonetheless, teachers, counselors, and even pupils and their parents have always made predictions about future performance, and they will no doubt continue to do so. The very decision to enroll in a course like Regents Chemistry which is elective and which is known to be intellectually demanding involves an element of prediction: the supposition that the pupil will be successful in meeting the course requirements. The purpose of statistical prediction, then, is to objectify prediction and, hopefully, to make it more precise and accurate than the subjective implicit predictions upon which important decisions are generally founded.

Making predictions about the future performance of groups of students is much safer, of course, than making predictions about individuals. In group predictions, substantial variations on the part of individuals balance one another, and the mean predicted score will be quite close to the actual mean achieved score. But systematic errors will still be reflected in the outcomes, so that differences in the difficulty level of the criterion test, differences in the students with respect to factors that are not considered in making the predictions but still exert an influence upon the criterion outcome, and differences that may not otherwise be identifiable all contribute to error.

In considering the results that are presented in this chapter, we must remind the reader that he is viewing the results associated with students who were enrolled in Regents Chemistry during the 1963-64 school year, and that the results apply, directly, only to the Regents Examination administered in June of that year. The accuracy of predictions based upon this experience, therefore, will be directly related to the congruence between the characteristics of pupils enrolled in chemistry during that year and in subsequent years, and to the equivalence of Regents examinations administered that year and in subsequent years.

Simple Correlation

Table XX gives the equations for the regression line connecting achievement scores on the Chemistry Pretest and each of the six unit tests

TABLE XX

REGRESSION EQUATIONS BETWEEN THE CHEMISTRY PRETEST AND THE UNIT TESTS AND THE 1964 CHEMISTRY REGENTS EXAMINATION

The Regents Examination score equals:

1.59 Chemistry Pretest score + 38.95 (r = .57; SE_{est} = 13.39)

1.51 Periodic Table score + 23.56 (r = .72; SE_{est} = 11.25)

1.49 Water, Solutions, and Ionization score + 27.79 (r = .72; SE_{est} = 11.37)

1.54 Halogen score + 23.47 (r = .76; SE_{est} = 10.64)

1.36 Sulfur-Nitrogen score + 28.70 (r = .70; SE_{est} = 11.62)

1.57 Organic Chemistry score + 21.27 (r = .77; SE_{est} = 10.48)

1.41 Principles of Reaction score + 26.49 (r = .81; SE_{est} = 9.67) with the scores on the Regents Examination. We note that the correlation coefficient is substantial between the Pretest and the Regents Examination and that the correlation coefficient increases markedly for the test on the first unit of instruction, Periodic Table and then increases more gradually for the tests given later in the school year. The standard error of estimate changes accordingly; the standard error of estimate from the Chemistry Pretest is substantially larger than the subsequent ones, and those decrease as the date of test administration approaches the date of the Regents Examination.

The information contained in Table XX can be given in another form which is easier to interpret. One can take the score achieved by a pupil on the Chemistry Pretest or on one of the unit tests and estimate his most likely score on the Regents Examination. This is done in Table XXI.

One enters Table XXI with the score the pupil achieved on one of the tests in the column headed by the name of the test. For instance, if a pupil scored 26 on the Chemistry Pretest, one would look in the column headed by "CHEMISTRY PRETEST" to the entry "26." Then, reading horizontally across to the column headed "REGENTS EXAMINATION," one obtains an estimate of his most likely Regents Examination score, an 80 in this hypothetical case. However, this is the most likely score only. Actual scores for pupils with a 26 on the Chemistry Pretest distribute themselves about 80. To get an estimate of how widely they will tend to range, we look at the last column entry under Chemistry Pretest, the 50 per cent range entry. This tells us that 50 per cent of the pupils who score 26 on the Chemistry Pretest have Regents Examination grades within 9 units of 80; 50 per cent have Regents Examination grades between 71 and 89. This hypothetical pupil then has a probability of .50 of scoring between 71 and 89 on the Regents Examination, a probability of

TABLE XXI

REGENTS EXAMINATION SCORES ASSOCIATED WITH SPECIFIC TEST SCORES

CHEMISTRY PRETEST	PERIODIC TABLE	WATER SOLUTIONS IONIZATION	HALOGENS	SULFUR NITROGEN	ORGANIC CHEMISTRY	PRINCIPAL REACTION	REGENTS EXAM.
38	50	49	50	--	50	--	100
35	46	45	46	49	47	49	95
32	43	42	43	45	44	45	90
29	40	38	40	41	41	41	85
26	37	35	37	38	37	38	80
23	33	32	33	34	34	34	75
20	30	28	30	30	31	31	70
16	27	25	27	27	28	27	65
13	24	22	24	23	25	24	60
10	20	18	20	19	21	20	55
9	8	8	7	8	7	6	FIFTY PER CENT RANGE

.25 of scoring below 71 on the Regents Examination, and a similar probability of .25 of scoring higher than 89.

Or, as another example, a student scores 27 on the Halogen test. Reading across from the entry in the "HALOGEN" column predicts 65 as the most likely Regents Examination score. For this pupil the data indicate a 50 per cent chance of passing that examination. With a 50 per cent range of 7 for the distribution of predicted Regents Examination, furthermore, the probability is .50 that his score will be between 58 and 72, .25 that it will be below 58, and .25 that it will be above 72. In constructing Table XXI, numbers were rounded to the nearest whole number. For intermediate scores, one must interpolate in both columns.

Multiple Correlation

Since approximately half of the total variance on most of the pairs of predictor tests appeared to be shared variance, the prospects for improved prediction through multivariate analysis seemed favorable. This was born out for pairs of predictor variables (Table XXII). The multiple correlation coefficients are larger and the standard errors of estimate are reduced. Employment of additional predictor variables beyond two, however, did not increase correlation sufficiently to justify the increased computational complexity. The table gives the multiple regression equations for successive pairs of tests used in this study. As before, a substantial part of the variance on the Regents Examination scores can be attributed to the variances of the earlier tests and the common variance increases as the dates of test administration approach the date when the Regents Examination is given.

The information contained in the multiple regression equations can also

TABLE XXII

MULTIPLE REGRESSION EQUATIONS BETWEEN PREDICTOR SCORES AND THE 1964 CHEMISTRY REGENTS EXAMINATION

The Regents Examination score equals:

0.69 Pretest score + 1.25 Periodic Table score + 17.96 (r = .75; SE_{est} = 10.72)

0.92 Periodic Table score + 0.85 Water, Solutions, and Ionization score + 17.64 (r = .78; SE_{est} = 10.23)

0.79 Periodic Table score + 1.00 Halogen score + 15.38 (r = .80; SE_{est} = 9.75)

1.06 Halogen score + 0.67 Sulfur-Nitrogen score + 17.10 (r = .80; SE_{est} = 9.82)

0.60 Sulfur-Nitrogen score + 1.11 Organic Chemistry score + 16.69 (r = .80; SE_{est} = 9.87)

0.76 Organic Chemistry score + 0.92 Principles of Reaction score + 17.52 (r = .84; SE_{est} = 8.81) be given in somewhat simpler form. For this purpose, we have constructed the nomographs in Figures 1-6. The nomographs are used as follows:

1. Locate the two test scores to be used in prediction on the appropriate coordinate axes. (Interpolation may be necessary.)
2. Find the point of intersection of these two coordinate in the nomograph.
3. Interpolate between the diagonal lines to estimate the most likely Regents Examination score. (The numbers on the diagonal are the most likely Regents Examination scores.)

For example, a student scores 35 on the Chemistry Pretest and 30 on the Periodic Table Test. In Figure 1, the intersection of those coordinate lines is very near the diagonal which predicts an 80 on the Regents Examination. However, we feel we must stress again that 80 is only the most likely score. A group of students with 35 on the Pretest and 30 on the Periodic Table Test will obtain Regents Examination scores that distribute themselves around 80 so that any one student has a good chance of scoring higher or lower than 80. In order to simplify the nomographs, ranges were not explicitly given. They can be computed from the multiple regression equations. However, all pupils whose predicted scores are at least 70 have a high probability of passing the course.

Since one of the chief uses of the nomographs may be in the early identification of pupils who are likely to fail, the three bottom left diagonals of the nomographs contain estimates of the probability of passing the course. Thus a student with a score of 30 on the Periodic Table Test and a score of 25 on the Halogen Test (Figure 3) has only a 50 per cent chance of passing the Regents Examination; with scores of 20 on both tests, his chance is slightly less than 10 per cent. This information may be helpful in changing study habits or in early recognition of probable failure.

We would suggest, however, that if the information of this chapter is

Figure 1

Nomograph for Regents Examination Scores Associated with Pretests and Periodic Table Test Scores

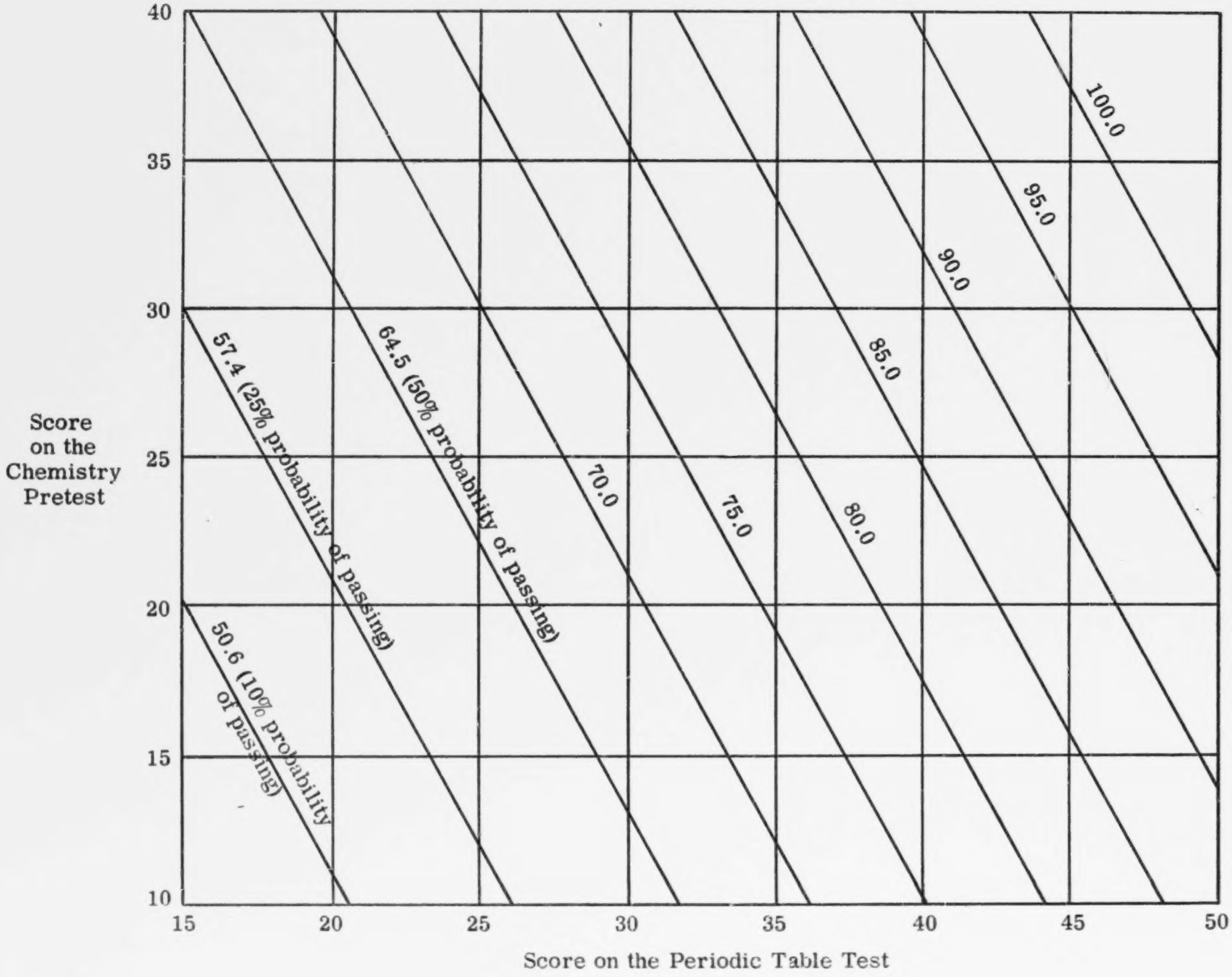


Figure 2

Nomograph for Regents Examination Scores Associated with Periodic Table and Water, Solutions and Ionization Test Scores

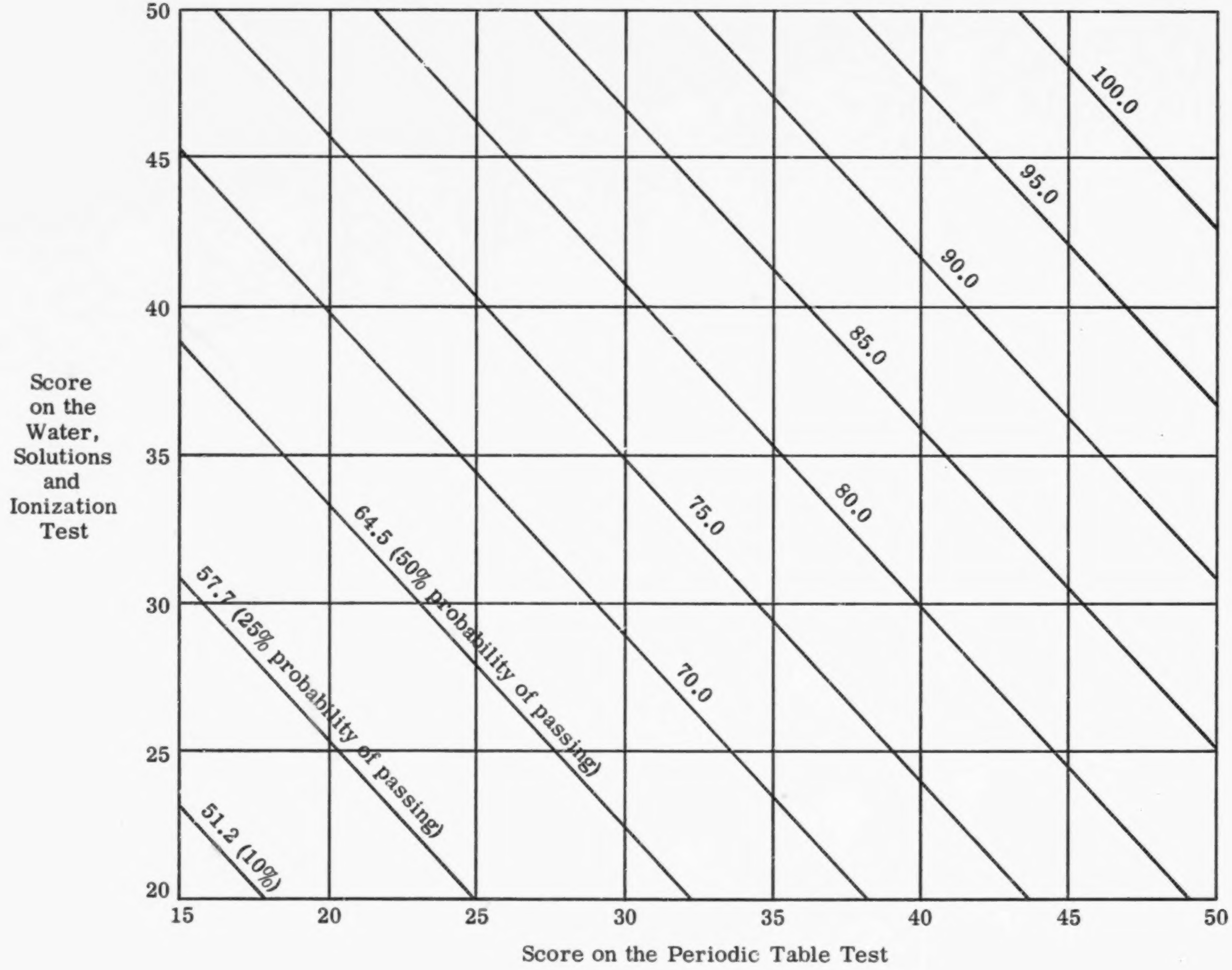


Figure 3

Nomograph for Regents Examination Scores Associated with Periodic Table and Halogen Test Scores

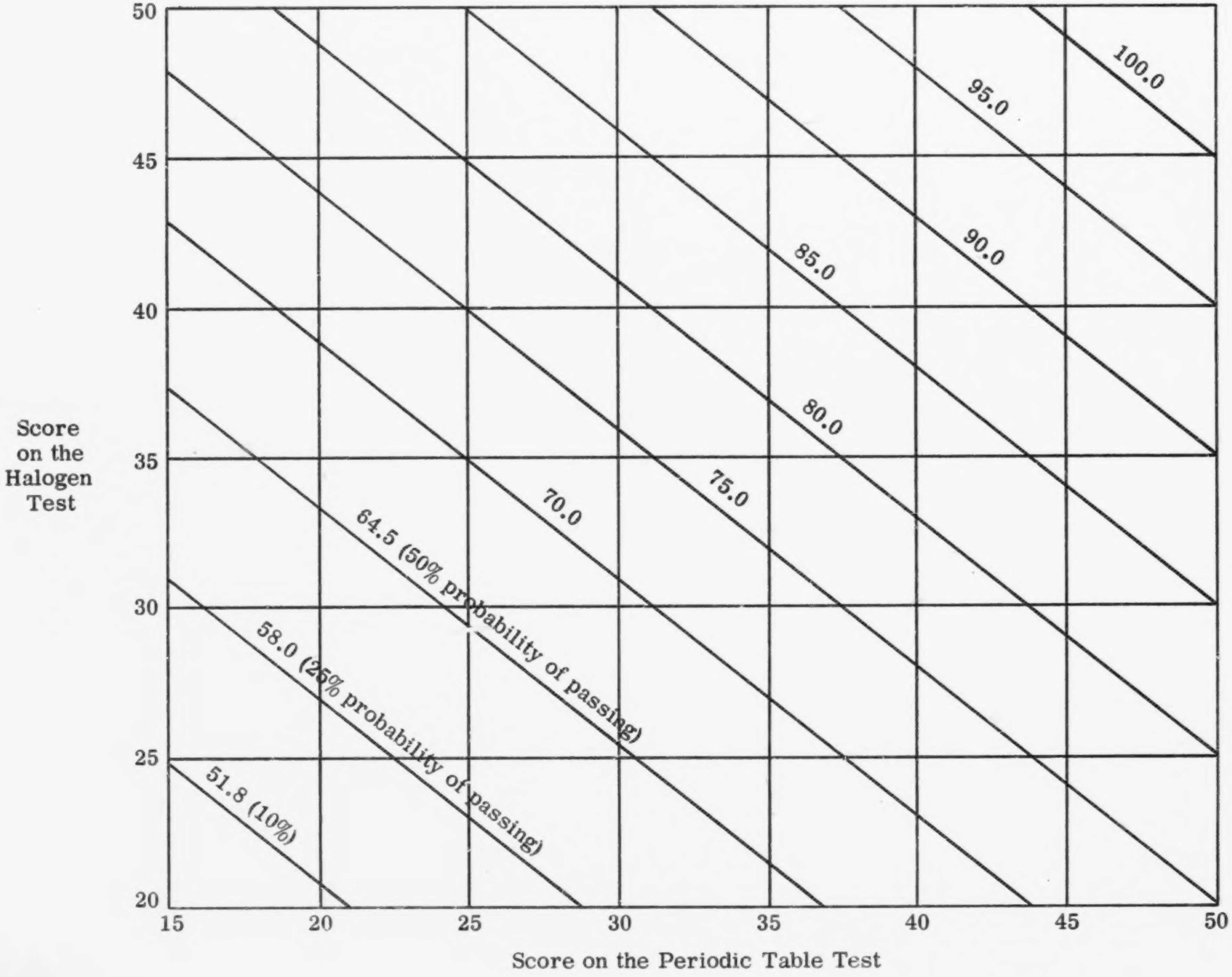


Figure 4

Nomograph for Regents Examination Scores Associated with Halogen and Sulfur-Nitrogen Test Scores

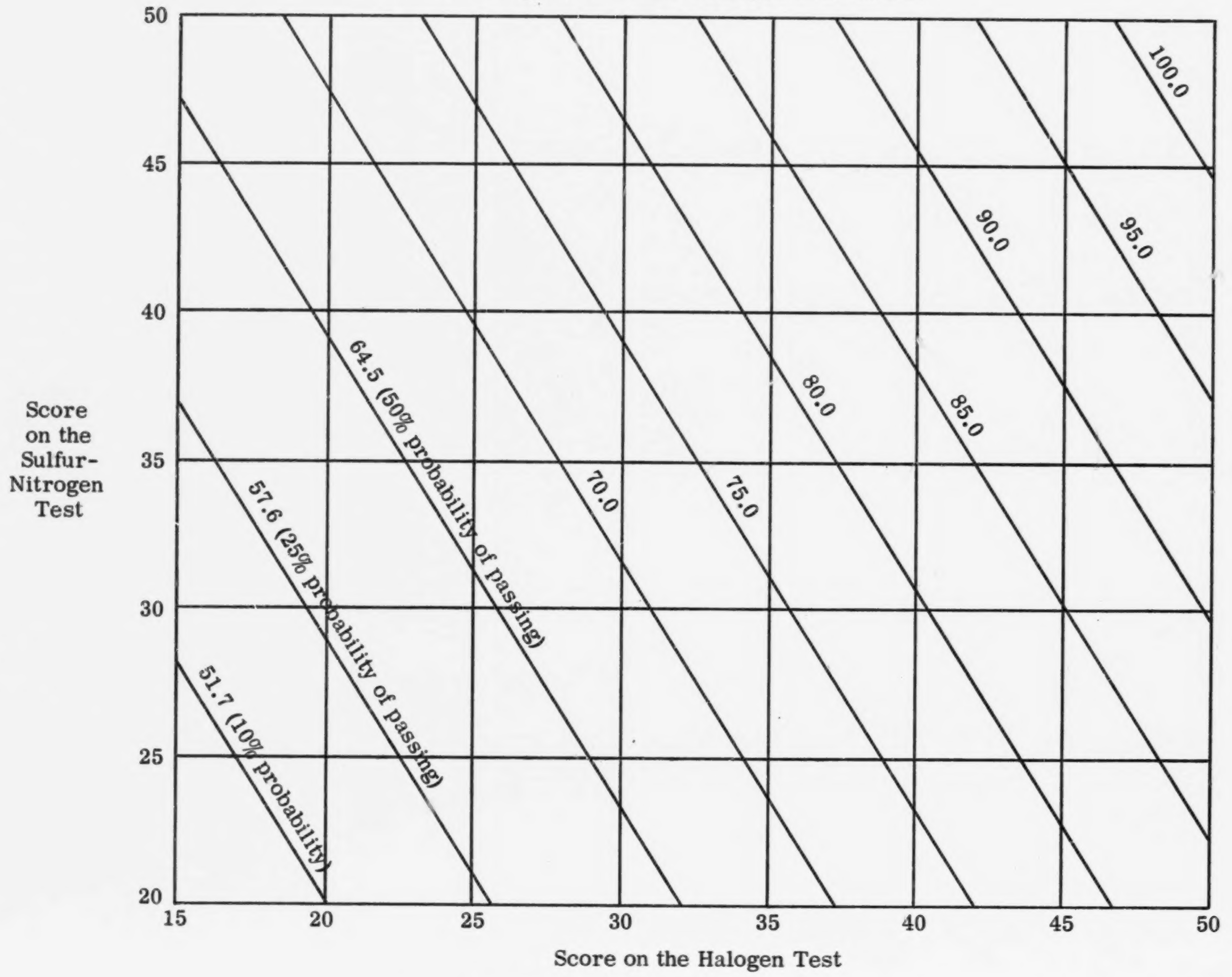


Figure 5

Nomograph for Regents Examination Scores Associated with Sulfur-Nitrogen and Organic Chemistry Test Scores

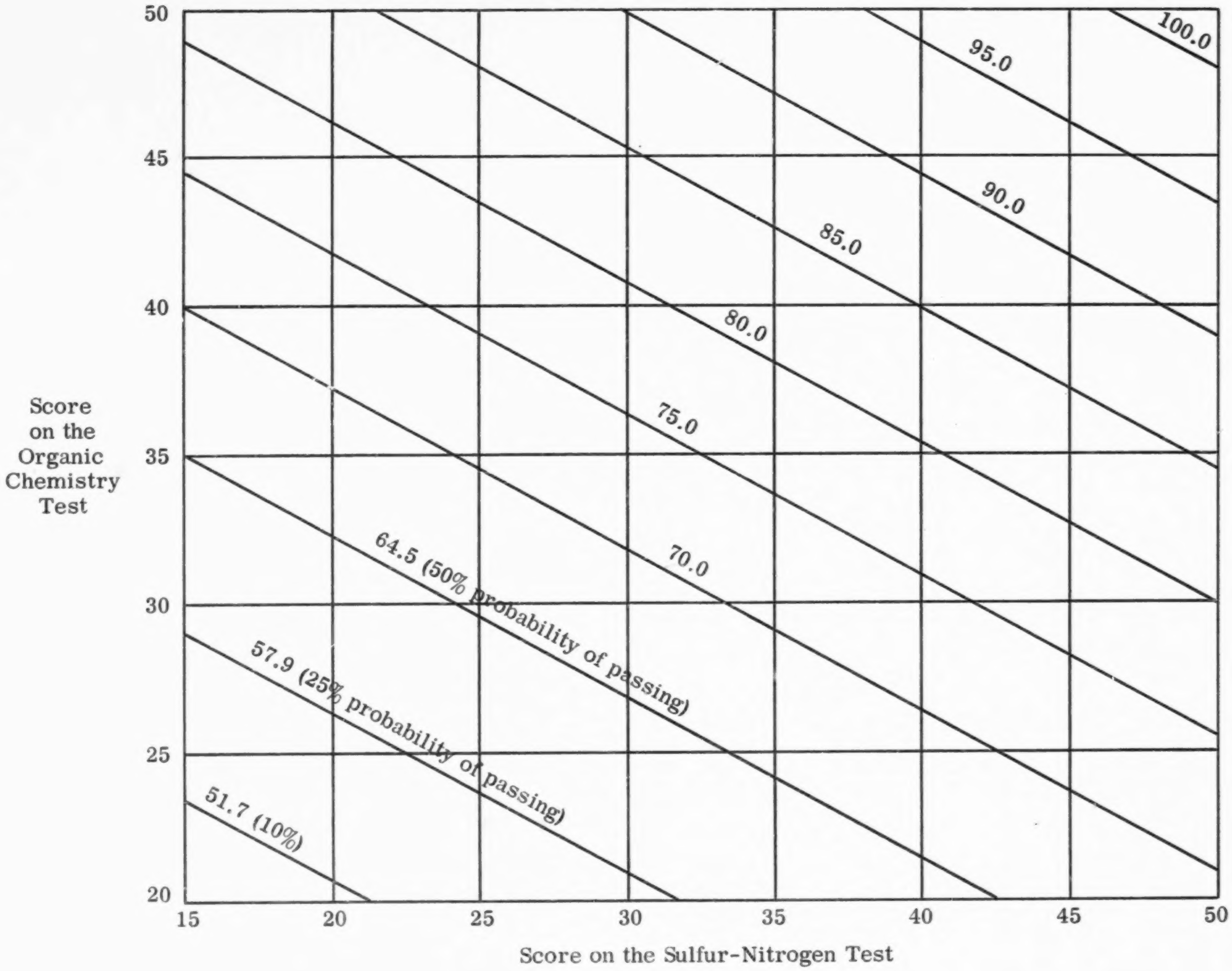
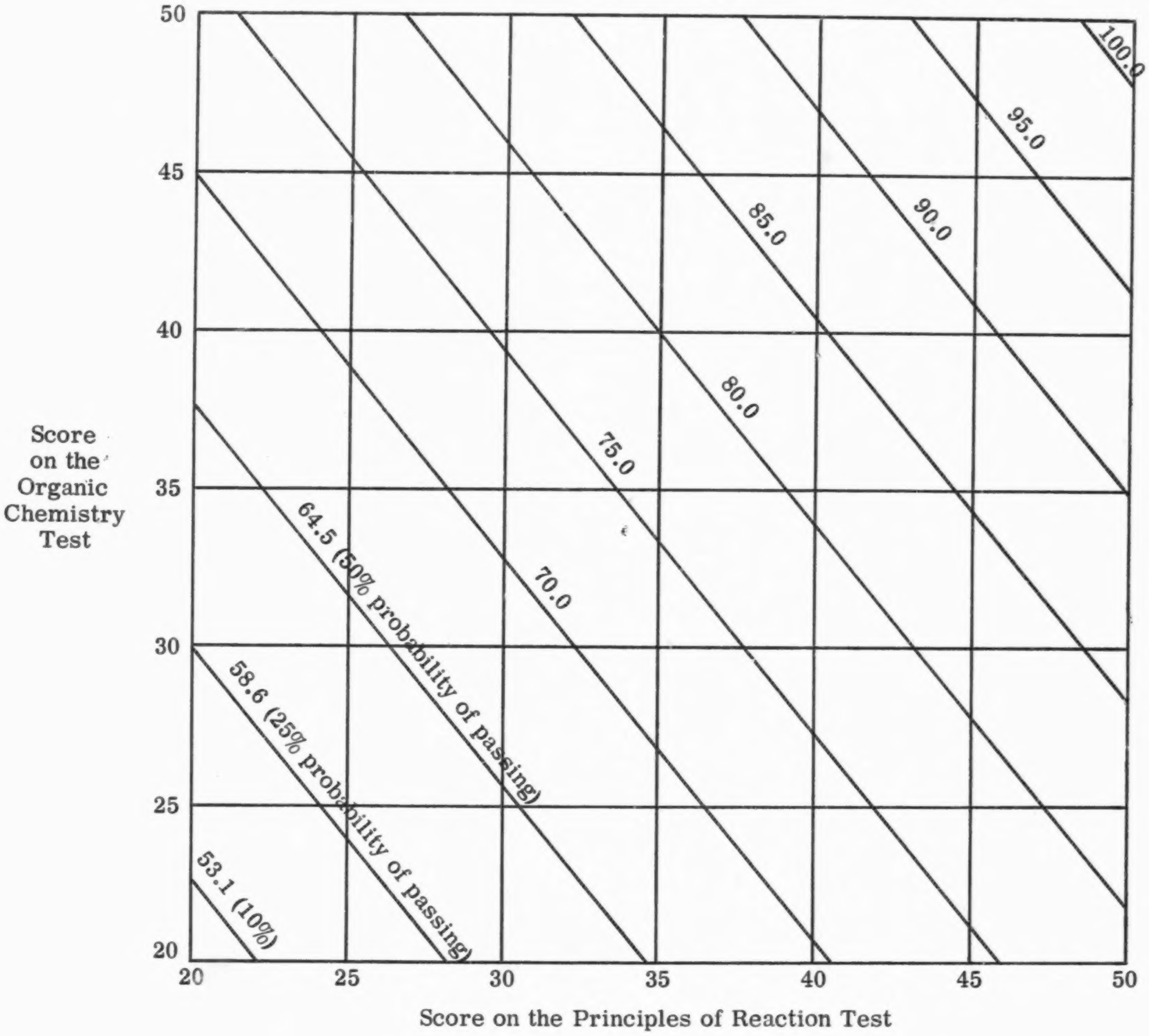


Figure 6

Nomograph for Regents Examination Scores Associated with Organic Chemistry and Principles of Reaction Test Scores



shown to students, that it be used without direct reference to prediction since the prediction is accurate only with groups of students, not with any single case. It is probably preferable simply to stress the experiences of other students who had similar scores on the predictor tests, and to point out that a number of other factors contribute to the determination of any particular student's Regents Examination score. It would probably be appropriate also to point out that the Regents Examination measures something somewhat different from what the predictor tests attempted to measure.

A final word of caution is in order with respect to the possible use of these regression data in districts other than those which participated in the study. To the extent that other districts, their teachers and/or their pupils differ from those represented in the study, predictions based on these data may be even less accurate than within the participant districts; there is no way of telling this without empirical evidence.

Chapter 6

INSTRUCTIONAL PROCESSES

With the large and significant differences on most achievement measures between large group - small group students and conventional students, it is particularly important to determine whether organizational pattern alone produces the different results or whether more specific teacher behaviors can be found to which achievement difference can be attributed. Consequently, we carried out several different analyses of the data from the Instructional Process Description Instruments in an attempt to discover commonalities among the practices of the large group teachers and of the conventional teachers. The details of the organizational patterns in the six large group schools (Table V) differed to such an extent that the mere existence of large groups and of small groups appeared as the only common feature among them.

The first analysis of variance attempted to determine differences in the time spent by teachers in the two patterns on each of 23 different teacher, pupil, or teacher-pupil interaction behaviors. These were the 23 behavior categories of the Instrument (Table XXIII). On none of the behaviors was there a significant difference which could be ascribed to organizational pattern. On most, variations within each organizational pattern were at least as great as the difference between large group - small group and conventional teachers. No distinct teacher behavior category could be identified which might be considered characteristic of one or the other of the organizational patterns.

An analysis of variance was also made on the ten objective categories

TABLE XXIII

BEHAVIOR CATEGORIES

Teacher does . . .

Teacher talked, while the students listened and/or took notes or recorded statements verbatim; illustration, demonstration, student questions, etc., were only incidentally employed if at all (I).

Teacher used chemistry-laboratory materials and/or equipment to illustrate some principle directly related to a specific objective of the course (I).

Teacher used other means of communication to transmit, demonstrate or illustrate facts or principles (e.g., films, filmstrips, recorders) (I).

Teacher administered a written test, the students having at least one day's notice or expecting the test on that day (II).

Teacher administered a written test, the students having less than a day's notice (II).

The teacher discussed with and/or lectured to the class on some topic having only incidental (or no) direct relation to any specific objective of the chemistry course (e.g., segregation, the four-minute mile, federal aid for education) (I).

Teacher and Student . . .

Teacher directed questions and designated individual students of his own selection to provide the answers orally. Students may or may not have previously seen questions (e.g., homework) (III).

Teacher directed questions but restricted his selection of the student to answer to those indicating their willingness (viz., by raising hand, etc.) (III).

Students directed questions to the teacher, and he answered them or elicited the answers from other students (IV).

Teacher and the students engaged in open interchange of ideas without formalized structure (i.e., students did not withhold comment or participation until recognized by the teacher) (IV).

Student . . .

Students engaged in open interchange of ideas without formalized structure, and the teacher's participation was minimal, or he took no part at all in the discussion (IV).

TABLE XXIII (Continued)

Individual or small group of students presented a demonstration, using chemistry-laboratory materials and/or equipment while the other students listened and/or took notes (III).

Individual or small group of students presented a report or panel discussion while other students listened and/or took notes. The presentation was primarily verbal; illustration and demonstration were used only incidentally if at all (III).

Students engaged in laboratory work (actual manipulation of lab equipment) directed toward the fulfillment of a specific course objective. Teacher gave normal supervision (V).

All students present worked at the blackboard solving problems assigned by the teacher (III).

Identical with above, except that only a part of the class worked at the blackboard, the remaining students observing and/or solving problems at their seats (III).

Students solved problems or wrote verbal responses to questions assigned by the teacher. Teacher exercised general supervision and gave advice or assistance as requested. (This may or may not be a head start on homework assignment.) (III)

Students studied text or other materials for general learning rather than to answer specific questions. Teacher exercised general supervision and gave advice when requested (VI).

Students scored their own or another student's test paper, under teacher's supervision (VII).

Students policed the laboratory or performed other tasks more directly related to maintenance of the classroom facility than to the achievement of specific course objectives (VII).

Miscellaneous . . .

The teacher engaged in some activity of a primarily procedural nature (e.g., taking roll, reading announcements, handing our papers, announcing homework or tests) (VII).

Interruption in the class (e.g., P-A announcement, fire drill--objective code X will usually be appropriate here) (VII).

Something not classifiable under any of the above categories (VII).

(Table XXIV) of the Instrument. These were designed to elicit the purposes of the instructional effort. Like on the behavior categories, no distinction was found which would divide the 15 teachers clearly into the large group - small group or conventional pattern. Time spent on various objectives of teaching also varied as much within each pattern as between the patterns.

Since no characteristics could be identified which clearly distinguished all large group teachers from all conventional teachers when teacher behavior categories were examined for each teacher individually, we grouped data to see whether any distinctive patterns emerge for either organizational pattern. Since many of the 23 behavior categories require only negligible fractions of the teaching time, this procedure failed to give an insightful description of the range of teaching behaviors, and the data are not reproduced here. Not surprisingly, conventional teachers spent three times as much time in directing questions for voluntary pupil response (15% vs. 5% of the teaching time). Also not surprisingly, large group teachers employed common audio-visual aids--films, filmstrips, etc.--twice as much as conventional teachers (8% vs. 4% of the teaching time). But perhaps surprisingly, demonstration-lecture time differed only slightly between the two patterns, with conventional teachers spending 27% of their time on this activity contrasted with 32% of time for large group teachers. Our initial hypothesis that the details of teacher behavior may differ less between the two patterns and that class size may be the major operational variable seems to be supported.

An expected pattern emerges when the 23 behavior categories are further grouped into groups of related interactions. By combining the categories as indicated by the Roman numerals in Table XXIII, we arrived at seven groups of behaviors: teacher dominance (I); evaluation (II); teacher initiated interaction

TABLE XXIV

OBJECTIVE CATEGORIES

1. To introduce and/or develop a concept relevant to the chemistry courses. This may involve specific facts, generalizations, reasoning procedures, definitions, names of equipment or descriptions of how equipment operates. This category should be used when the primary reason for carrying out the activity is the initial development of a new concept, even though references may be made to concepts previously developed in order to further the development of the new concept.

2. To correct an erroneous concept acquired by one or more students in the chemistry class, in some other class, outside of classes, or prior to beginning the course in chemistry. This category should be used only when some positive effort is made to correct; in some instances erroneous concepts will be eliminated as the students gain an acquaintance with new concepts, and in such cases Code 1 would be more appropriate, since the elimination of the erroneous concept would really be incidental to the primary goal of developing a new concept.

3. To review concepts previously developed for the primary purpose of increasing retention. This would be an objective only after the initial development of the concept (as represented by Code 1) had been completed. In some instances, tests may be given for this purpose rather than to evaluate concept development, and in cases where a test is given for purposes of evaluation as well as to provide overlearning experiences, the reporting teacher will have to base his selection of the appropriate code number on his personal decision regarding the relative priority of the two objectives.

4. To develop a motor or visual-motor skill to a degree of proficiency that makes possible its practical application. This category is appropriate for all procedures for initial introduction to the skill to the stage where errorless performance first becomes possible. Correction of performance errors is included in this category.

5. To provide practice in the application of a skill for the express purpose of enhancing the probability of its retention. (This category should not be used when utilization of the skill is merely incidental to performance of a laboratory exercise which has concept development as its primary goal.)

6. To develop an attitude which is clearly a part of the chemistry-course objectives. This category would be appropriate when efforts are directed toward the development of a personal respect for scientific objectivity; it is also appropriate when time is spent in attempts to motivate (i.e., to create a positive attitude in) pupils toward chemistry.

7. To develop an attitude that is not directly related to the objectives of the

TABLE XXIV (Continued)

course in chemistry. Time might conceivably be spent in discussing attitudes toward minority groups, or in attempting to motivate one or more pupils with respect to school work in general, in which case this category would be more appropriate.

8. To evaluate the progress of one or more (or all) pupils in concept development. The primary purpose here should be evaluation of progress; if increased learning takes precedence, category 1, 2, or 3 should be used.

9. To evaluate the progress of one or more (or all) pupils in skill development. The primary purpose here should be evaluation of progress; if improved skill performance or enhanced retention of a given level of performance is more important than the evaluation process in the teacher's planning for the procedure being reported, then category 4 or category 5 should be used.

10. The objective of the reported activity was something different from any of the above. (If this category is used, a brief statement of the appropriate objective of the activity in question should be made in the remarks section or on the reverse side of the reporting form.)

(III); student initiated interaction (IV); laboratory work (V); supervised study (VI); and miscellaneous and administrative work (VII). When these are compared for the two organizational patterns, the expected greater time spent in teacher dominated activities in the large group classes comes out clearly (Table XXV). Large group teachers spend more than 43% of the time lecturing, demonstrating, showing films, etc. Conventional teachers, however, also spend much time in such activities, 33.5% of the time. (Modern ideas of individualizing instruction have not reached chemistry classes.) Teacher directed interactions take up the next largest time fraction, 33% of the conventional class time, 23% of the large group class time. With time spent on evaluation, these two teacher dominated activities take up 80% of school time in both kinds of schools.

On activities in which the student takes the initiative only slight differences were noted. Large group students spent about 12.5% of their time in student initiated interactions whereas only 10.7% of the time of the conventional classrooms was available for these activities. Large group students spent about 5% of their time in supervised study against about 6% for conventional students. Perhaps of some importance is the fact that more than 4% of conventional class time was devoted to administrative and miscellaneous chores against less than 2% among the large group schools. It should be noted that our sampling procedure tended to give inadequate weight to the laboratory, and that the negligible time reported for it is a product of this factor.

Our data on teacher behaviors confirm certain suspicions but do not give conclusive evidence. They confirm the notion that the personal characteristics of the teacher are more significant in determining his style of teaching than the organizational pattern in which he finds himself. They confirm that a major fraction of the time is spent in teacher dominated behaviors in

TABLE XXV

COMPARISON OF GROUPED TEACHER BEHAVIORS

BEHAVIORS	PER CENT OF TIME LARGE GROUP	 CONVENTIONAL
Teacher Dominance	43.5	33.6
Evaluation	13.8	12.2
Teacher Initiated Interaction	22.8	33.0
Student Initiated Interaction	12.5	10.7
Supervised Study	5.5	6.2
Miscellaneous and Administration	1.8	4.3

conventional classrooms as well as in large group classes. They confirm that a sizeable amount of class time is spent on administrative and miscellaneous routines. The data also indicate that large group students have slightly better access to the teacher for student initiated behaviors, no doubt in the small group sessions, and consequently spend somewhat more time getting their problems and difficulties discussed. However, aside from these slight, inconclusive indications, we did not learn a great deal about the teacher behaviors under the two organizational patterns. Certainly we did not isolate a reason, or even obtain an intimation of a reason, for the consistently greater achievement among students in the large group - small group organization. Unless we accept the extra few per cent of time spent in student initiated discussion as the reason for higher achievement, direct teaching behaviors apparently did not influence the results of the study.

This leads to two hypotheses concerning our findings. Either our data are inadequate and we found no differences in the instructional processes because our instrument was inadequate to the task, or the differences between large group - small group and conventional classes result from the pattern of organization alone and not from the instructional behaviors under it. We are tempted to accept both hypotheses. We feel that our instrument for recording teacher behaviors was not adequate to the needs of the study. Its failure is unfortunate although not surprising. Researchers have labored for at least 50 years on the task of adequately defining teacher behaviors with experiences and results similar to ours. In any case, we judge that our instrument was not incisive enough to elicit small differential factors between the two organizational patterns which might have produced the achievement differences. But we also feel that factors related to the organizational pattern alone and independent of teacher characteristics helped produce the results. The change from the tedious routine of the academic classroom of 25 pupils, 5 times a week, we suggest, produces a "Hawthorne Effect" in the pupils; the demands for individual responsibility on the part of the pupil in the large group lecture, we feel, leads to somewhat more independent study which carries over into tested achievement. Certainly the latter suggestion receives some corroboration from the greater drop-out rate in large group - small group schools in which, we think, more dependent pupils find inadequate support to sustain their efforts. In summary, our instrument did not show major differences in the instructional processes that could be attributed exclusively to organizational pattern; however, we believe that the instrument was not adequate to elicit such differences should they exist. On the other hand, we propose that the instructional behaviors of teachers in the large group - small group schools are not the sole cause for the achievement differences we observed and that factors such as the novelty and relative impersonality of the large group class contribute to produce them. Finally, we want to point to the large differences in preparation time which are discussed in the next chapter. No doubt this influences the effectiveness of instruction.

Chapter 7

COST

Educational decisions are never made in terms of a single factor. Educational value, cost, availability and readiness of staff, community factors, and a number of other considerations influence them. Hence, our attempt to assign specific costs to chemistry instruction, or rather, to determine whether there were cost differentials between instruction under the large group - small group pattern and the conventional pattern constituted a significant part of this study. But while cost would appear to be the most readily measurable aspect of instruction, it proved quite difficult to pinpoint to an instructional area. School accounting does not assign specific costs to specific courses or departments. Consequently, we had to develop a procedure for estimating costs related to instruction in Regents Chemistry.

In actual fact, we were unable to estimate the cost of instruction in chemistry in any single school. The best we were able to achieve was a reasonable estimate of differentials between schools using the two patterns of organization. In order of decreasing amounts involved, such differentials depend on differences in the effectiveness of the use of staff, differences in the space needs for the two patterns of instruction and the possibility of integrating the chemistry space with other instructional space requirements, differences in the need for capital equipment, and differences in the use of non-capital instructional supplies. Each of these was considered separately in our study. Other cost factors were judged too small and too diffuse and were assumed inconsequential.

Despite a serious attempt, the last item, supplies, could not be isolated for comparison. Expendable supplies are typically ordered for the entire school or the entire department. Even when items are specifically purchased for chemistry courses, they are typically bought in quantities large enough for several years, and in a period of expanding enrollments, time averages are nearly impossible to compute. We could not isolate the relevant data. However, we could see no *a priori* reason for any cost differentials since both organizational patterns were generally similar in their manner of carrying out instruction and laboratory work. Also, supplies are a minor fraction of the total budget. Thus, while we were forced to this decision by unavailability of data, we feel that we are justified in assuming that any differentials in supplies costs between the two organizational patterns are negligibly small.

Among the capital items, we found no substantial differences. Most schools in the large group organization had purchased overhead projectors, but so did many conventional schools. These were used for the chemistry instruction and for other instruction in both kinds of schools. The expected period of amortization was long so that the annual cost factor became negligibly small compared to other cost factors. In capital equipment items, also, we could not isolate costs specifically for chemistry instruction and we saw no evidence for assuming that appreciable differences existed between the schools in the two instructional patterns.

Space utilization and space cost can be determined quite accurately. Building costs are on record and space use can easily be measured in square feet and hours. We attempted an analysis of this kind but found that local variables made our data meaningless. Building costs, quite obviously, depend on the year of construction. One school in the study was opened in 1963; another had built a science wing two years earlier. Other buildings were built in the 50's, 40's, 30's, and 20's. Some buildings contained additions built at various times. This made comparisons useless. But much more difficult to analyse are differences quite far removed from the cost of the chemistry space. An incisive study of space costs would have required far more time and specialized skill than was available for this study.

The space devoted to science depends on many factors. Some schools use separate classroom and laboratory spaces. Some have combined spaces. The most recently constructed schools, and one older one, had amphitheater-type classrooms specifically designed for large groups. Two used auditoria; one adapted a room for this purpose. We were forced to conclude that taste, availability of space, integration into the overall program, and similar non-measurable factors determine the space use in a school and that in our limited study, it was impossible to assign specific dollar values to either instructional pattern. Schools with relatively ample space for science instruction and schools with relatively little space existed among both kinds.

Since building costs are a large budgetary factor, we could not assume that differentials are negligible although amortization of school plant reduces this item also. But we can not express any differentials in more precise terms than that space costs seem to reflect the overall philosophy of the school no matter what organization is employed for chemistry instruction.

Only in the area of direct instructional costs and instructional overhead costs could we proceed with our analysis to the point of obtaining meaningful data for comparisons. On instructional overhead--supervision and planning--we found that only two large group schools expended funds for planning, and these were of the order of $300 spent only once, prior to setting up the program. No school identified specific administrative or supervisory costs that might differentiate large group instruction from conventional instruction. While we presume that some additional supervisory work was required in setting up large group instruction in all six schools, it obviously was incorporated into the normal supervisory and administrative routines. Administrators and supervisors, we suspect, simply did not differentiate time spent on setting up large group instruction from time budgeted for normal program development. This leaves the largest budget item, instructional salaries, as the sole means to differentiate costs between the two instructional patterns.

As we indicated earlier in this report, an adjusted pupil-teacher load rather than salary was used as the basis for teaching cost comparisons because of the extraneous items that determine the exact compensation received by a particular chemistry teacher. Salary schedule and fringe benefits differ from district to district and in each case reflect the advanced education and experience of the teacher. Yet the concern of a comparison of instructional costs is with whether more or fewer teachers are required under one pattern or the other rather than whether the teacher happens to be high on the salary schedule and happens to have advanced degrees or education.

The adjusted pupil-teacher load was computed from the fraction of the class periods actually spent in chemistry instruction, the fraction of assigned preparation periods presumably devoted to chemistry preparations, and the number of pupils who were instructed in chemistry. Adjustments were made for the total number of instructional periods per week in the school and the number of periods per week each pupil spends in chemistry instruction. An illustration of the computation is given by an example:

Pupils:	Number of pupils instructed in Regents Chemistry	78
	Number of periods per week spent in Regents Chemistry	6
Teachers:	Number of periods assigned to Regents Chemistry	15
	Number of periods assigned to other instruction	5
	Number of periods assigned, non-instruction (study hall, etc.)	3
	Number of periods assigned to preparation	12
	Total per week	35

The teacher in this school spends 15 periods teaching Regents Chemistry. If we assume that he divides his 12 preparation periods proportionately between chemistry and his other teaching assignment of 5 periods, three-fourths of the 12 preparation periods must also be charged to chemistry. Hence, he devotes 15 plus 9 or 24 hours to Regents Chemistry. This means that he devotes 24/35 or .685 of the school week to chemistry. At the same rate of instruction, if he devoted the entire school week to chemistry, he could instruct 78/.685 = 114 pupils in the 6 period per week chemistry course. If the course met only 5 periods per week, he could instruct 1.2 times as many pupils. Adjusted for this factor also, the pupil-teacher load becomes 137 pupils per week. (This last adjustment was needed to reduce 5, 6, and 7 period per week chemistry courses to a common denominator.)

Computations of the type illustrated here were made for all large group schools (supplemented by one school not in the study which also instructs under the large group - small group plan) and for all conventional schools in the study (supplemented by 12 other schools). Data for the additional schools came from returns received to a questionnaire sent to a random selection of Western New York high schools. Table XXVI shows the results. The adjusted

TABLE XXVI

ADJUSTED TEACHER-PUPIL LOADS AND TEACHER PREPARATION TIME

	CONVENTIONAL	LARGE GROUP
Number of Schools	20	7
Number of pupils taking Regents Chemistry		
Range	18-340	85-195
Median	63	130
Range, Adjusted Pupil-Teacher Load	75-170	82-204
Range, Preparation Periods Per Week	2-11	5-15
Range, Percent of Time Available for Preparation (Preparation Periods/Total Periods Per Week)	6-28%	18-43%
Mean Adjusted Pupil-Teacher Load	122*	130*
Standard Deviation	21.1	39.1
Mean Preparation Periods	6.1**	10.8**
Standard Deviation	2.1	3.2
Mean Percent of Time Available for Preparation	16.5%**	32.3%**
Standard Deviation	5.2	7.5

*Not significant at 0.5 level

**Difference significant at 0.0005 level pupil-teacher loads do not differ significantly between the large group schools and the conventional schools. Neither organizational pattern is more or less expensive to staff than the other because teachers under both instruct just about the same number of pupils. However, in preparation time, the large group teachers fare significantly better. Nearly one-third of their working time is available for preparations (which may be one important factor in the higher achievement of their pupils) in contrast to one-sixth of the time for conventional teachers. While the range of preparation time overlaps in the table because one of the seven large group teachers had much less preparation time than his colleagues, all remaining large group teachers had more preparation periods than any conventional teacher. (It should be emphasized that the tests of significance are merely indicative, not confirmatory, because of sampling problems.)

Our analysis of costs under the two instructional patterns leads to the following conclusions. There is no difference in the cost of instruction under the large group - small group or the conventional pattern. Cost of space seems to be determined by factors other than organizational pattern. Teachers under the large group - small group organization have substantially more time for preparation without diminishing their service in terms of numbers of pupils. Cost is not a factor which should influence toward or away from large group - small group instruction.

Chapter 8

RECOMMENDATIONS

The evidence gathered in this study is relatively self-explanatory. On 13 achievement measures corrected for initial pupil differences, pupils taught under the large group - small group instructional patterns surpassed pupils taught conventionally to statistically significant extents. This finding was applicable to students who initially scored high on the control measures as well as to pupils who scored low on these measures. It applied to boys as well as girls (with some exceptions). On measures of interest, no significant differences between large group - small group taught pupils and conventionally taught pupils were found. Also, no cost differences were found between the two organizational patterns while large group teachers had additional preparation time. Hence, it appears that large group - small group instruction is an educationally desirable pattern of instruction.

Nevertheless, we can not, on the basis of this study, recommend that all schools adopt large group - small group instruction without pointing very specifically to a number of limitations of this study which reduce the general applicability of its findings.

Sampling Problems

The first limitation of our results is the nature of the sample. Schools which participated in this study did so on their own initiative. They agreed to support the study financially which implies a research orientation among their teachers, administrative staffs, and boards. They were primarily suburban schools. The assumption of random sampling of schools in the study is definitely not tenable.

Secondly, the large group - small group schools had decided on that pattern of instruction independent of the study. Their teachers, administrators, and boards were therefore distinguished from those in the conventional schools by their prior independent decision to try large group instruction.

Thirdly, all teachers in the schools which participated had been identified by the State Education Department as having substantial experience with Regents Chemistry and having in the past had classes who did well on the Regents Examination. Hence, the findings are definitely limited to experienced teachers who have successfully taught the Regents Syllabus.

Because of the restrictions of the sample, any application of this study is, in our view, limited to schools in which parallel conditions hold. We feel that the achievement results obtained here have a high probability of being duplicated only in suburban schools with experienced and successful teachers after the staff, administration, and boards have reached the decision, on other bases, to try large group - small group instruction.

Presumed Success Factors

A second set of cautions relates to our inability to define precisely differences in behavior between teachers in the two organizational patterns. This forces us to assume that the organizational pattern, in and of itself, was the determining factor in the observed differences. Consequently we must hypothesize about the importance of the large group - small group itself.

All the large group schools used that pattern of organization in courses other than chemistry, but in no school was the large group - small group used extensively. Most schools used it only for one grade level of history instruction. Hence, the findings, it seems to us, are dependent on the fact that the large group - small group pupils found the pattern somewhat of a novelty. We hypothesize that this novelty was important (the Hawthorne Effect). We would be seriously concerned if our findings were generalized to other subjects with the result that students experienced few or no conventionally taught classes in the later high school years. Entirely different results might then occur.

Secondly, we feel that the extra preparation time of teachers in the large group - small group pattern has a causal relation to the higher achievement. We strongly suspect that erosion of the extra preparation time by administrative decision, after the newness of this organizational pattern fades and it loses its experimental designation, would produce different results. We found educational value to large group - small group instruction, not lower costs. We fear that an attempt to add saving as a second value, by reducing the preparation time, would lead to the sacrifice of its educational values.

Finally, we assume that a reason for the higher achievement in large group - small group instruction was the fact that students could not depend as much on their teachers for instruction but had to rely on their own studies to a greater extent. This is, of course, a double value: greater achievement and greater independence on the part of the student. Yet, this factor had its negative aspects as well. Some students withdrew from chemistry, presumably, because they were unable to make the adjustment. These students constitute a loss of undetermined magnitude. Regents Chemistry is normally taught in 11th grade. Pupils in this age group can be expected to show considerable maturity, yet they are still late adolescents. We do not know whether the withdrawals were of students who learned their limitations early rather than later, in college perhaps, or of students who needed additional time to mature into independence. We suspect both groups were represented and are concerned about the loss of the latter.

Recommendations

In view of these cautions, our recommendations are conditional. If a school staff, administration, and school board reaches the decision to instruct by the large group - small group organizational pattern, and it is in a suburban area and has an experienced and successful chemistry teacher, there is a high probability that achievement in the Regents Chemistry course will increase and that the instruction will cost no more than under the conventional pattern. However, the achievement gains will probably be accompanied by a greater withdrawal rate from chemistry. The gains will be limited to academic achievement and will not include an increase in interest or a differential gain in science reasoning and understanding. Moreover, we feel that these gains are predicated on the fact that the large group teacher has substantially more time available for preparation. If these conditions are met, we definitely feel that we can recommend the decision to adopt large group - small group instruction as an educationally sound decision.

Chapter 9

A NOTE ON PROCEDURE

The research described in this report is not an experiment in the classical sense of the term. We did not randomly choose and assign treatments and in several other respects we violated the norms of experimental design. It was an applied, not a theoretical investigation and we are well aware of its limited character. Our study clearly reflects the complex factors that are involved in the improvement of education through research (Brickell, 1961). Clearly we operated in the realm of application rather than in the realm of fundamental investigation. The character of the problem, however, demanded such an approach.

Large group - small group instruction is a phenomenon of considerable current interest. It is being applied in schools for a number of reasons although, as we indicated earlier, without adequate knowledge of its true functioning. The first questions to be asked about it, it seems to us, is not why does it work or even how does it work, which would be questions that would require detailed experimental investigation, but rather does it work? And it was to the latter question that we devoted all efforts.

From the purely practical point of view, from the point of view of either encouraging or discouraging school administrators in the use of this pattern, the latter question seems to us by far the most important. It would be redundant to spend three years on a detailed investigation of how large group instruction works only to find out, at the end, that it is detrimental to the student. Obviously a school administrator wants to know first that it works before he would try it in his school.

It seems to us, likewise, that our research would have limited generality and applicability in schools if our procedures had demanded a set pattern of behaviors among the large group teachers. From a theoretical point of view, it is of course extremely desirable to know precisely what kinds of behaviors are likely to lead to optimum results. Then, teacher training efforts can be directed to developing these behaviors. (Although this goal has been pursued in research for at least 50 years, it has produced extremely meager results. We still do not know the behavior of a good teacher and indeed we do not even know how to define him.) But to demand that large group teachers who wish to duplicate our results change their teaching practices to a specific set of behavior patterns, or even to train experimental teachers in these patterns when we do not know which behaviors are likely to be effective, seemed to us an impractical procedure. In view of these circumstances, we felt that a restriction on the performance patterns of our large group - small group teachers would limit the usefulness of our findings rather than increase our understandings.

The application of our results involves at least three sets of probabilities all of which must be fairly high or our study is extremely restricted in its usefulness. The first set of probabilities relates to the statistical significance of our results. Here we found that the probabilities are extremely large that schools similar to those in our study would find increased achievement through large group instruction. The next set of probabilities relates to finding schools similar to those in our study. In this regard our sample was restricted to suburban schools with experienced and successful chemistry teachers. The probability that our findings will be replicated is therefore restricted to similar schools. This fits a substantial fraction of the schools in New York State or in the nation and consequently the probability of finding similar schools is fairly high. The next probabilities relate to the instructional behaviors of the experienced and successful teacher in those comparable schools. Because of the design of our study, this probability is extremely high also because we placed very few, if any, restrictions on the teacher behaviors. Both the specific school arrangements for large group instruction--time and spaces provided--and the teacher behaviors were left entirely to the discretion of the particular school and the particular teacher. Consequently we can feel that most arrangements for teaching large groups of pupils, with associated small group and laboratory meetings, and most individual styles for teaching which do not deviate extremely far from the more or less customary patterns represented in our study would meet this criterion at an extremely high level. Thus we feel that our design maximized all three sets of probabilities governing the reproduceability of our results and, as a consequence, produces findings of fairly high generality and usefulness.

We designed our study in order to provide maximum information of a practical sort to school administrators desiring to try large group - small group instruction in their schools. This design forced us into an empirical procedure that violated some of the canons of basic research. It precluded our testing of certain hypotheses regarding fundamental relationships under large group - small group pattern and, for that matter, of contributing much to basic knowledge regarding the learning process. On the other hand, it gives some relatively clean cut answers to the practical question: "What is the likelihood that some specific suburban school would benefit from the adoption of large group-small group instruction in chemistry?" The administration of that school must judge to what extent teachers and school are like the teachers and schools of our study. If the similarities are large, they can have considerable faith that pupils in their school, too, will attain higher achievement under large group-small group instruction.

REFERENCES

Bloom, B. S., "Thought Processes in Lecture and Discussion," Journal of General Education, Volume 7, 1953.

Bloom, B. S., editor, Taxonomy of Educational Objectives, New York, David McKay Co., 1956.

Breedlove, Charles B., "An Appraisal of Team Teaching in High School Chemistry," Dissertation, Wayne State University, 1963.

Brickell, Henry M., Organizing New York State for Educational Change, Albany, State Education Department, 1961.

Flanders, Ned A., "Interaction Analysis in the Classroom: A Manual for Observers," Unpublished manuscript, University of Michigan, 1960.

Fleming, C. M., "Class Size as a Variable in the Teaching Situation," Educational Research, February 1958.

Gerberich, J. R., and Warner, K. C., "Relative Instructional Efficiencies of the Lecture and Discussion Methods in a University Course in American National Government," Journal of Educational Research, Volume 29, April 1936.

Guilford, Joy P., "Three Faces of Intellect," American Psychologist, Volume 14, 1959.

Gulliksen, Harold, Theory of Mental Tests, New York, John Wiley and Sons, Inc., 1950.

Johnson, R. H., and Lobb, M. D., "Jefferson County College, Three Year Study," Bulletin of the National Association of Secondary School Principals, Volume 45, January 1961.

Lisonbee, Lorenzo, "Large Group Teaching," The Science Teacher, Volume 29, February 1962.

Ruja, H., "Outcomes of Lecture and Discussion Procedures in Three College Courses," Journal of Experimental Education, Volume 22, June 1954.

Trump, J. Lloyd, and Baynham, Dorsey, Guide to Better Schools, Chicago, Rand McNally & Company, 1961.

Watson, Fletcher G., and Cooley, William W., in Rethinking Science Education, Fifty-ninth Yearbook of the National Society for the Study of Education, Chicago, University of Chicago Press, 1960.